Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Animal

Study on the diversification of flat periwinkles (*Littorina fabalis* and *L. obtusata*): insights from genetics and geometric morphometrics

João Gonçalo Monteiro Carvalho

Dissertação

Mestrado em Biologia Evolutiva e do Desenvolvimento

2014

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Animal

Study on the diversification of flat periwinkles (*Littorina fabalis* and *L. obtusata*): insights from genetics and geometric morphometrics

João Gonçalo Monteiro Carvalho

Dissertação

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada pela Professora Drª. Manuela Coelho e

Dr. Juan Galindo Dasilva

2014

Acknowledgments

Bem, demorou mas finalmente se conclui este capítulo da minha vida. Ao longo destes meses foram várias as pessoas com quem partilhei este caminho mas parece-me adequado começar pelas pessoas que me acompanharam em Vigo.

Suponho que seja melhor escrever esta parte no meu espanhol limitado, por isso aqui vai: Sin un orden en particular, me gustaría comenzar haciendo referencia a Celeste y Eva por ayudarme en los primeros días en Vigo, cuando no conocía a nadie y no podía hacer un gel! Muchísimas gracias también a las chicas del café (Raquel, Ana, Nieves, Sara) porque siempre estáis alegres y muy amables! Para MJ, gracias por la compañía en las mazmorras y también por la música. A todas las personas en el grupo (Ángel, Andrés, Antonio, Armando, Emilio y Paloma) gracias porque siempre estabais amables conmigo y por crearen un ambiente de trabajo divertido. Tengo que mencionar el grupo de Comeeeeer!!! Por la compañía y las conversaciones humorísticas e interesantes durante la comida y el café posterior, gracias Lara, Dani (Nos vemos en Lisboa) y Belén! Por último, un agradecimiento muy especial a Mónica por cada momento de diversión, por todas las conversaciones y por aguantar con mis tonterías.

Hablar de Vigo es imposible sin hablar de Juan Galindo (XB2, Universidad de Vigo). También sería imposible enumerar todas las razones que tengo para darte las gracias. Desde tu paciencia con la burocracia de las universidades hasta tu paciencia con mis errores, pasando por todo lo que me has enseñado durante este tiempo. Gracias por todo lo que he aprendido contigo y por tu amabilidad y apoyo constante! Martín es un tío con suerte :)

Voltando agora ao idioma de Camões e Fernando Pessoa, vou aproveitar que falei do Juan para falar sobre o meu outro orientador, o Rui Faria (CIBIO, Universidade do Porto). Quero agradecer-te por me integrares na equipa do fascinante projecto de investigação que lideras, financiado pela FCT e FEDER ("The paths of parallel evolution and their genetic crossroads" - PTDC/BIA-EVF/113805/2009), e que permitiu que esta tese fosse uma realidade. Quero também agradecer-te por tudo o

que me ensinaste, tanto na parte de análise de dados como na parte da escrita. Esta tese não teria metade da qualidade que tem sem as tuas correcções e melhorias. Obrigado também por todo o teu apoio e por acreditares em mim e principalmente por tudo o que me ensinaste sobre especiação em particular e sobre ciência em geral. Neste ultimo ponto, uma das coisas mais importantes foi perceber que apesar de ser difícil gerir e coordenar um grupo, é possível fazer isso mesmo e ao mesmo tempo criar um ambiente positivo onde várias questões científicas podem ser abordadas e debatidas.

Não posso falar do Rui sem referir as pessoas que trabalham neste grupo fantástico. Graciela, muchas gracias por toda tu ayuda con los micros y muchas otras cosas. Eres una persona incansable y espero que nunca cambies pues eres genial! Obrigado também a Diana pela ajuda com aquela base de dados infernal com a morfometria. Neste último aspecto, tenho também que agradecer ainda pela ajuda e sugestões feitas, ainda que de uma forma indirecta, pela Antigoni Kaliontzopoulou (CIBIO).

Tenho também que agradecer a todas as pessoas que participaram e ajudaram bastante na parte de amostragem deste trabalho: Ao Rui e ao Juan primeiro que tudo, mas também à Diana Costa, à Graciela Sotelo, à Teresa Muiños Lago, à Carolina Pereira, ao Petri Kemppainen (University of Manchester), ao Mårten Duveport e a Kerstin Johannesson (ambos da University of Gothenburg). Dentro deste excelente grupo de pessoas, gostava ainda de destacar a ajuda da Carolina Pereira com as extracções de DNA de amostras do Norte da Europa e a ajuda da Teresa Lago por me ensinar tudo o que sei sobre o pénis de caracóis (sexagem) e fotografia dos indivíduos. Uma última menção para a professora Susana Varela que me pôs em contacto com o Rui Faria e, desta forma, tornou possível a minha participação neste projecto e para a minha orientadora interna, a professora Dr.ª Manuela Coelho por estar sempre disponível e se preocupar comigo.

Uma referência especial para a Sara Rocha por ser como é e por me deixar ficar em sua casa tantas vezes! Obrigado por seres sempre simpática e deixa andar (como

um verdadeiro português deve ser!) e por fazeres o derradeiro sacrifício de dormir na sala alguns dias quando eu estava por aí!

De um ponto de vista mais pessoal, quero agradecer aos amigos que tornam tudo mais fácil e mais agradável! Aos amigos da faculdade com quem sempre se pode contar para conversas animadas (ph, Jonas, Mariana, Miguel e Ana), um muito obrigado! Quero também agradecer aos meus amigos que fizeram com que nunca me sentisse demasiado longe de Portugal e do Benfica! Ao Daniel, ao Francisco e ao Mário, um muito obrigado! A ti, Leonor, um obrigado especial pela companhia que me fizeste nos momentos em que me sentia mais sozinho em Vigo. Graças a ti as saudades apertaram menos e apenas se mantiveram como algo saudável e familiar.

Por ultimo, um obrigado do tamanho do mundo para os meus pais, os meus irmãos e os meus avôs. Obrigado por serem me apoiarem e se preocuparem comigo! Obrigado pelos telefonemas, pelas conversas, pelas visitas e por tudo o resto que não se pode traduzir em palavras! No fundo, obrigado por sempre estarem aí para mim! Gosto muito de vocês todos! :)

Resumo

Especiação, a evolução de isolamento reproductivo entre diferentes populações, tem sido um dos temas mais investigados na área da biologia evolutiva. Dado que os investigadores têm geralmente acesso a apenas um instante deste processo contínuo, o estudo de especiação coloca, por definição, enormes desafios metodológicos. Um modo de ultrapassar estas dificuldades é através do estudo das diferentes fases de divergência dentro do mesmo sistema, desde os estádios inicias de diferenciação até ao isolamento reproductivo completo. Esta comparação pode-nos informar sobre os mecanismos e as forças que actuam nos diferentes estádios, assim como o modo como interagem durante o "contínuo da especiação".

Quando uma espécie ocupa diferentes habitats, a acção da selecção natural divergente pode resultar em importantes barreiras ao fluxo génico entre populações localmente adaptadas e possivelmente originar novas espécies, num processo geralmente denominado de especiação ecológica. A zona intertidal é dos ambientes mais heterogéneos do planeta, sendo caracterizada por gradientes abruptos de determinados factores abióticos e bióticos que podem variar numa escala de poucos metros. Neste contexto, a acção da selecção natural divergente em populações que habitam a zona intertidal pode levar à formação de ecótipos e até mesmo de novas espécies.

As espécies do género *Littorina* (gastrópodes marinhos que vivem na zona intertidal), para as quais vários ecótipos têm sido descritos, são cada vez mais reconhecidas como excelentes modelos para estudar as causas ecológicas de especiação. Duas destas espécies, *Littorina fabalis* e *L. obtusata*, são o alvo deste trabalho. Apesar de inúmeras semelhanças, estas espécies-irmãs diferem em factores como a localização na zona intertidal (*L. fabalis* vive em zonas mais expostas à ondulação do que *L. obtusata*), o tamanho dos indivíduos, e principalmente ao nível da morfologia do pénis, sendo esta última a principal característica utilizada na distinção entre ambas.

Em *L. fabalis*, diferentes ecótipos foram identificados desde o extremo Norte da distribuição da espécie (Suécia, Noruega e Reino Unido) até ao Sul, na Península

Ibérica. No Norte da Europa, dois ecótipos foram classificados de acordo com o seu tamanho e o nível de exposição as ondas: o ecótipo LM (*Large Moderatly exposed*) apresenta um maior tamanho e uma maior exposição à força das ondas; enquanto os indivíduos do ecótipo SS (*Small Sheltered*) são menores e habitam áreas mais abrigadas. Na Península Ibérica, foram descritos três ecótipos, que além de diferirem no grau de exposição às ondas, encontram-se normalmente associados a distintas algas: o ecótipo ME (*Mastocarpus Exposed*) consiste em indivíduos de pequeno tamanho que habitam algas do género *Mastocarpus* em zonas altamente expostas; em zonas de exposição intermédia, encontra-se o ecótipo FI (*Fucus Intermediate*) composto por indivíduos de maior tamanho do que os ME e que habitam algas do género *Fucus*; e por fim, o ecótipo ZS (*Zostera Sheltered*) com indivíduos ligeiramente maiores que os FI, que foi apenas encontrado numa localidade bastante abrigada na Galiza, vivendo em ervas marinhas do género *Zostera*.

Neste trabalho foram caracterizadas genética e morfologicamente várias populações de ambas as espécies, incluindo todos os ecótipos descritos para *L. fabalis* em duas grandes zonas geográficas (Península Ibérica e Norte da Europa), com o principal objectivo de entender os mecanismos envolvidos na diversificação (formação de ecótipos e espécies) deste grupo. Para tal, 13 populações de *L. fabalis* e uma população de *L. obtusata* repartidas por vários pontos da Galiza e do Norte de Portugal foram amostradas e analisadas para uma bateria de microssatélites que se desenvolveu especificamente para estas espécies. Adicionalmente, populações de três países do Norte da Europa (Suécia, Noruega e Reino Unido) foram também alvo de estudo. Em cada um destes países, um ponto exposto e um ponto abrigado, característicos de cada um dos ecótipos (LM e SS, respectivamente) foram amostrados em pelo menos duas localidades, e os indivíduos analisados com base em AFLPs (*Amplified Fragment Lenght Polymorphism*).

Tanto as populações da Península Ibérica como do Norte da Europa foram estudadas de um ponto de vista morfológico através do método de morfometria geométrica. Este método permitiu identificar diferenças relevantes na morfologia da concha entre *L. fabalis* e *L. obtusata*, de tamanho entre os vários ecótipos de *L.*

fabalis na Península Ibérica, e de tamanho e forma entre os dois ecótipos no Norte da Europa.

Na Península Ibérica, os resultados obtidos com 16 microssatélites confirmam a clara separação genética entre *L. fabalis* e *L. obtusata* e mostram a utilidade destes marcadores moleculares na discriminação das duas espécies, revelando pela primeira vez, a existência de hibridização entre elas. Ao nível dos ecótipos de *L. fabalis*, os resultados parecem indicar uma subestruturação genética das populações relacionada com a geografia, assim como uma maior diferenciação do ecótipo ME, provavelmente associada com uma acentuada deriva genética nestas populações mais expostas. No entanto, o papel da selecção natural na diferenciação destes ecótipos sugerido pela sua divergência fenotípica não foi ainda esclarecido, sendo para tal necessário o estudo futuro de marcadores moleculares envolvidos na adaptação aos diferentes habitats.

No Norte da Europa, o estudo de AFLPs detectou efeitos de selecção entre os indivíduos de zonas expostas e protegidas (LM e SS) em cerca de 5% do genoma, e uma percentagem de partilha de outliers relativamente elevada entre países e dentro de cada país (> 30%). A estrutura genética inferida através de loci *outlier* (selectivos) e *nonoutlier* (neutrais) revela, respectivamente, agrupamento por ecótipos e por geografia (Reino Unido vs. Escandinávia), sendo este padrão compatível com a formação dos ecótipos em paralelo e de um modo independente, em resposta à acção da selecção natural em cada uma destas áreas geográficas. A sequenciação futura dos outliers detectados neste trabalho, deverá esclarecer quais as possíveis origens (e.g. polimorfismo ancestral, fluxo genético, mutações de novo, ou uma combinação das três) da variação genética envolvida na formação e diversificação dos ecótipos.

O principal objectivo deste trabalho consistia em lançar as bases necessárias para tornar estas espécies num modelo de estudo de especiação ecológica. A combinação das ferramentas genéticas e morfométricas aqui desenvolvidas permitiu identificar duas fases distintas (um estádio inicial e outro mais avançado) de um processo contínuo de especiação ecológica: a formação de ecótipos de *L.*

fabalis no Norte da Europa e na Península Ibérica; e a diferenciação entre *L. fabalis* e *L. obtusata*. No entanto, a realização de estudos futuros será importante para clarificar quais os mecanismos e as forças responsáveis pela diversificação destes gastrópodes marinhos em particular, e de especiação ecológica em geral.

Palavras-chave: Littorina, ecótipo, divergência fenotípica, especiação ecológica, selecção natural.

Summary

Currently, speciation is viewed as a continuous process (i.e. the speciation continuum), where different mechanisms are involved in the buildup of barriers to gene flow across time. One of these mechanisms is natural selection, which is responsible for processes of ecological speciation. Here I study, phenotypically and genetically, the two sister species of flat periwinkles, *Littorina obtusata* and *L. fabalis*, including ecotypic variation of the latter associated with different habitats. Two main regions of their distribution were studied independently and with different molecular markers, the Iberian Peninsula (IP) and Northern Europe (NE).

In the IP, clear phenotypic (shell size and shape) and genetic differentiation was found between *L. fabalis* and *L. obtusata,* but also the first unequivocal evidence for hybridization between them, supporting the power of this approach for species discrimination and hybrid identification. Within *L. fabalis*, population (neutral) genetic structure better corresponds to geography than to ecology; whereas shell size divergence between ecotypes points to a possible role of natural selection in their diversification, although future studies are required to confirm these results.

In NE, a relatively high proportion of shared AFLP outlier loci was detected between *L. fabalis* ecotypes across countries in a genome scan for divergent selection. The genetic structure of the outlier loci showed ecotypes to group together over a large geographical scale, combined with their repeated phenotypic divergence, supports that LM and SS ecotypes are likely diverging under the influence of natural selection in the face of moderate gene flow; while the study of neutral loci (nonoutliers) showed a separation between Scandinavia and the UK, maybe related with the recolonization process after the last glacial maximum.

Thus, this work shows how different processes could be involved in the diversification of flat periwinkles across different stages of the speciation continuum, with natural selection likely acting as a key player; confirming the potential of this system to investigate parallel ecological speciation.

Key Words: flat periwinkles, ecological speciation, speciation continuum, phenotypic divergence, genome scan.

Index

Introduction

1. Natural selection and speciation

Speciation, i.e. the evolution of reproductive isolation between populations, has been one of the most debated topics in evolutionary biology (Coyne & Orr, 2004). The study of such a continuous and often long process is inherently challenging and further complicated by the fact that researchers used to have access to a single snapshot of it. Thus, the assessment of different stages of the evolution of reproductive barriers ("*the speciation continuum*", Hendry, 2009; Nosil et al., 2009a), representing different levels of divergence, is a powerful approach to overcome this main challenge and increase our knowledge on the mechanisms involved in speciation (Nosil, 2012).

Among these mechanisms, natural selection is perhaps the one that has received most attention. In the last decades, in particular, we have witnessed a renewed recognition of the role of ecologically-driven divergent selection as a major promoter of speciation (Nosil, 2012; Faria et al., 2014), which led to the emergence of the term "ecological speciation", defined as "*the process by which barriers to gene flow evolve between populations as a result of ecologically based divergent selection between environments*" (Nosil, 2012).

Ecological speciation is closely tied with the phenomenon of local adaptation (Butlin et al., 2014). Different environmental conditions in heterogeneous habitats pose divergent selective pressures, and adaptation to a particular micro-habitat (i.e. local adaptation) can lead to population divergence and, in certain cases, originate distinct ecotypes - "*spatially distinct populations exhibiting divergent adaptation to alternative environments*" (Funk, 2012).

One line of evidence for the role of natural selection in diversification is the repeated evolution of the same phenotypic traits and barriers to gene flow as a consequence of adaptation to similar environments in different localities (i.e. parallel speciation; Rundle et al., 2000; Schluter, 2000; Nosil, 2012). However,

demonstrating that ecologically-driven reproductive isolation between populations independently evolved in each locality involves not only finding evidence that phenotypic differentiation was caused by natural selection but also that reproductive isolation is the result of such differentiation (Faria et al., 2014).

In this respect, lower genetic distance between ecologically-divergent populations (e.g. ecotypes) within a locality compared to ecologically-similar populations from different localities is frequently interpreted as ongoing parallel ecological speciation (Nosil, 2012). Nonetheless, gene flow within each locality could reduce neutral genetic differentiation suggesting that they have evolved in parallel even if they did not (Faria et al., 2014). Thus, the use of neutral markers alone to distinguish these alternative scenarios may originate misleading interpretations.

Moreover, the analysis of genetic variation influenced by natural selection (e.g. detected with genome scans; Butlin, 2010) is expected to show reduced gene flow between ecotypes, although the origin of this variation can be substantially distinct (e.g. de novo mutations, gene flow between localities, standing genetic variation; Johannesson et al., 2010; Faria et al., 2014). Therefore, different cases of parallel speciation can show diverse patterns of genetic structure for loci under selection (e.g. outlier loci; Butlin, 2010) (Faria et al., 2014).

Examples of parallel speciation have been described in a wide range of organisms: the benthic and limnetic forms (Rundle et al., 2000; Boughman et al., 2005) and the freshwater and marine forms (McKinnon et al., 2004) of threespine sticklebacks, host-plant ecotypes of *Timema* walking stick insects (Nosil, 2007), exposed and sheltered ecotypes of *Littorina* marine snails (Johannesson et al., 2010; Butlin et al., 2014), among others. Despite the enormous progress in recent years, we have limited knowledge about the mechanisms responsible for speciation and how they interact, as well as about the genomic architecture of the process, highlighting the need for more studies across different stages of the *speciation continuum* to move the emerging field of speciation genomics even further (Seehausen et al., 2014).

2. Ecological speciation in flat periwinkles

The marine intertidal is a heterogeneous environment that comprises local abrupt gradients of several physical conditions (e.g. wave exposure, temperature) and biological interactions (e.g. predation) (Raffaelli & Hawkins, 1996). One well characterized example of ecological speciation through local adaptation at different levels of the intertidal zone is the case of the marine gastropod *Littorina saxatilis* (Rolán-Alvarez et al., 2007; Butlin et al., 2014).

Other *Littorina* species where phenotypic adaptive divergence and reproductive isolation have also been studied, although in less detail, are *L. obtusata* (Seeley, 1986; Trussel & Etter, 2001) and *L. fabalis* (Tatarenkov & Johannesson, 1998; Johannesson & Mikhailova, 2004). These two sister species, commonly referred as flat periwinkles due to their flattened whorled shell, are the focus of this study. They live in close association with macrophytes (e.g. fucoid algae), feeding on them (*L. fabalis* grazes microalgae from the surface of its host plant, while *L. obtusata* directly consumes the fucoid tissue) and with the females laying egg masses on the algae's thallus. The juveniles are formed before hatching, without an intermediate planktonic phase. These species, similarly to *L. saxatilis*, have poor dispersal capacities because of the lack of planktonic larvae (Reid, 1996).

The two species have a widely and largely overlapping distribution along the Northeast Atlantic (Reid, 1996) (Figure 1). However, only *L. obtusata* occurs in North America, resulting from the recolonization of the region from Europe after the last glaciation, about 13,000 years ago (Wares & Cunningham, 2001).

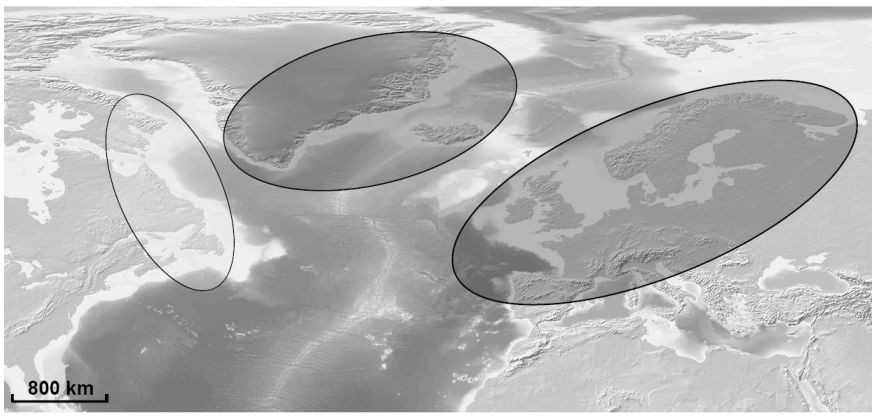

Figure 1. Distribution along the North Atlantic for *Littorina fabalis* and *L. obtusata*. While in Europe and Greenland the two species overlap (dark), in North America only *L. obtusata* is present (light). Information from Reid (1996).

Despite the similarities, some important differences do exist between *L. fabalis* and *L. obtusata* (Table 1), such as their preferential location along the intertidal shore, with *L. fabalis* individuals commonly found in more exposed zones when compared to *L. obtusata* (Reid, 1996). However, many of these characters are either of subjective interpretation, difficult assessment or are not completely diagnostic but rather represent a phenotypic continuum without clear boundaries between the two species. For example, *L. fabalis* individuals are generally smaller and possess a more flattened shell, but the intraspecific variation is so large that shell morphology is far from being a diagnostic trait. One exception is penis morphology (Table 1, Figure 2), which constitutes one of the most used (and reliable) traits to distinguish individuals from the two species (though it does not apply to females, which are still phenotypically difficult to classify).

Table 1. Distinctive characters between *Littorina obtusata* and *L. fabalis*

Character	*L. obtusata*	*L. fabalis*
Head-foot pigmentation	Darker	Paler
Paraspermatozoa	11-16 μM diameter	14-21 μM diameter
Penial filament	Short, triangular	Long, vermiform
Mamiliform glands	10-54 in a double row	3-17 in a single row
Copulatory bursa	Long	Short
Ovipositor	Usually pigmented	Usually unpigmented
Shape of egg mass	Oval or rarely kidney	Oval or often kidney
Mean ovum diameter	210-255 μM	195-200 μM
Radula	Usually 5 cusps	4 cusps
Adult size	Larger where sympatric	Smaller where sympatric
Relative columellar width	Thinner	Thicker
Sexual dimorphism	Slight	Pronounced
Wave exposure	Sheltered (i.e. upper shore)	Sheltered/moderately exposed (i.e. closer to the sea)

Adapted from Reid (1996)

On the other hand, genetic differentiation between the flat periwinkles has been observed at several allozyme loci (Rolán-Alvarez et al., 1995), indicating that genetic tools can provide crucial information for species identification. Indeed, recently, Kemppainen et al. (2009) were able to demonstrate a clear separation between *L. fabalis* and *L. obtusata* from Northern Europe (NE) based on microsatellite loci, although they also found a lack of clear differentiation for mitochondrial DNA (mtDNA). Nonetheless, the limitations of these datasets

compromise the ability to make genome-wide generalizations and to accurately characterize putative hybrids between the two species.

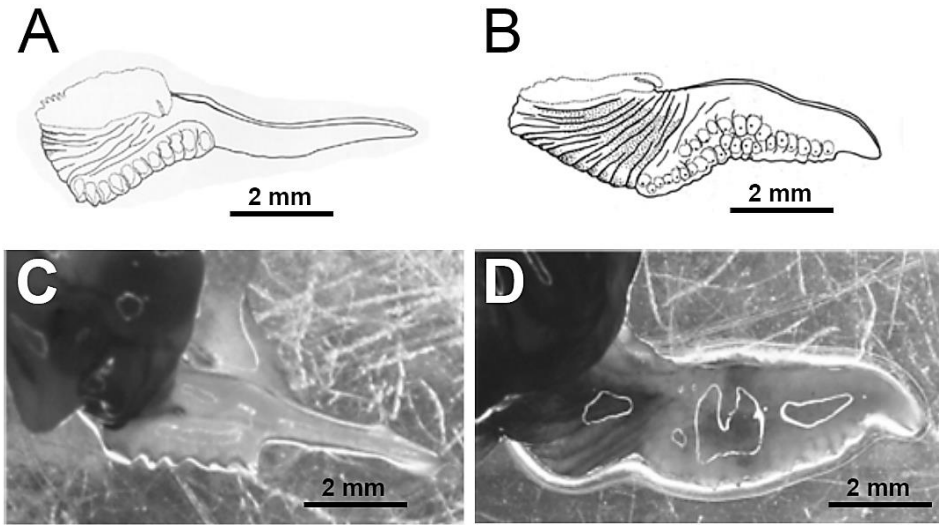

Figure 2. Typical penis morphology of *Littorina obtusata* (A and C) and *L. fabalis* (B and D). Note the difference in size of the penis filament and in number and arrangement of the mamiliform glands. Panels A and B are adapted from Reid (1996).

3. Intraspecific diversification of flat periwinkles

Similarly to *L. saxatilis*, where several ecotypes have been described in multiple geographic regions (Butlin et al., 2014), the flat periwinkles present phenotypic variation, and in certain cases genetic variation, associated with different microhabitats (Reid, 1996; Johannesson, 2003; Kemppainen et al., 2011).

In *L. obtusata*, significant variations in shell size and shape have been reported across its distribution range associated to different wave-exposure conditions (Reid, 1996). However, they have not been systematically studied. In contrast, several *L. fabalis* ecotypes occupying different habitats and facing different regimes of wave-exposure have been the targets of several studies. In NE, two locally adapted ecotypes were identified (Figure 3). While in shores of moderate wave-exposure, individuals are bigger and have a broader shell with a relatively larger aperture ('Large-Moderately exposed' ecotype, LM); in more sheltered habitats, shells are smaller and present a narrower aperture ('Small-Sheltered' ecotype, SS)

(Tatarenkov & Johannesson, 1999; Johannesson & Mikhailova, 2004; Kemppainen et al., 2005, 2009). Nevertheless, both ecotypes mostly dwell in the canopy of fucoid algae (mainly *Fucus* spp. and *Ascophyllum* spp.).

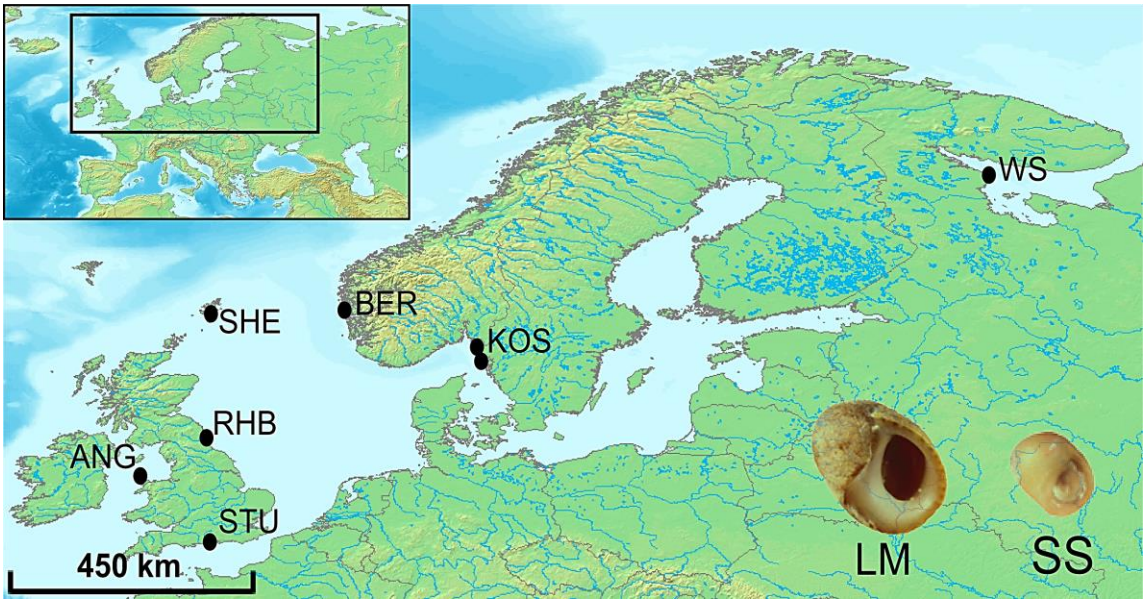

Figure 3. Large-Moderately Exposed (LM) and Small-Sheltered (SS) ecotypes of *Littorina fabalis* and their known distribution in Northern Europe. ANG – Anglesey; STU – Studland; RHB – Robinhoods Bay; SHE – South Shetland; BER – Bergen (Norway), KOS – Koster (Sweden - both ecotypes have been described in various off shore islands) and WS – White Sea (Kandalaksha) (detailed locations described in Tatarenkov & Johannesson (1994) and Kemppainen et al. (2009)).

In the Iberian Peninsula (IP), three different ecotypes were described (Figure 4; Rolán & Templado, 1987; Williams, 1990; Lejhall, 1998; Tatarenkov & Johannesson, 1998), not only associated with different levels of wave-exposure (as the Northern ecotypes), but also with different "host" algae, although the two factors are probably correlated. In areas of intermediate wave-exposure, the '*Fucus*-Intermediate' ecotype (FI) is characterized by medium-size snails associated with the brown algae *Fucus* spp.; the '*Zostera*-Sheltered' ecotype (ZS) is found in a single location occupying an extremely sheltered habitat, where snails of larger size inhabit the green seagrass *Zostera* spp.; and on the most extreme wave-exposure conditions that *L. fabalis* can tolerate, a dwarf and red/brownish ecotype is found associated with the red algae *Mastocarpus stellatus*, the '*Mastocarpus*-Exposed' ecotype (ME).

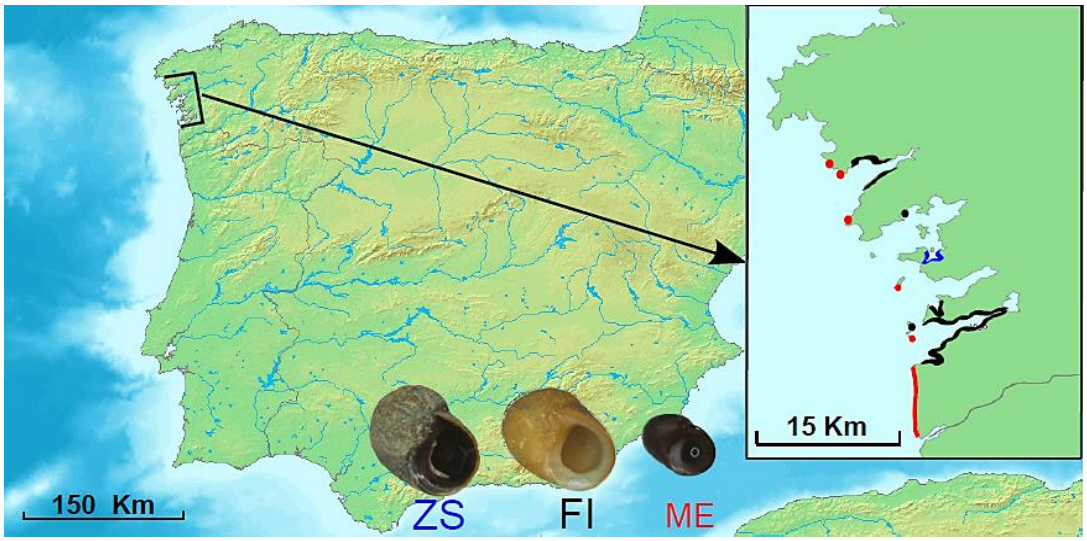

Figure 4. *Littorina fabalis* ecotypes and their distribution in the Iberian Peninsula, as available at the onset of this work. The distribution of the *'Zostera-Sheltered'* (ZS) ecotype is colored in blue; in black, of the '*Fucus*-Intermediate' (FI) ecotype; and in red, of the '*Mastocarpus*-Exposed' (ME) ecotype (detailed locations from Rólan-Alvarez & Templado (1987) and Rólan-Alvarez (1995)).

The association between shell and host-algae colors that is generally observed (Figure 5), probably reflects camouflage to avoid predators (Reimchen, 1979), although further studies are needed to properly test this hypothesis.

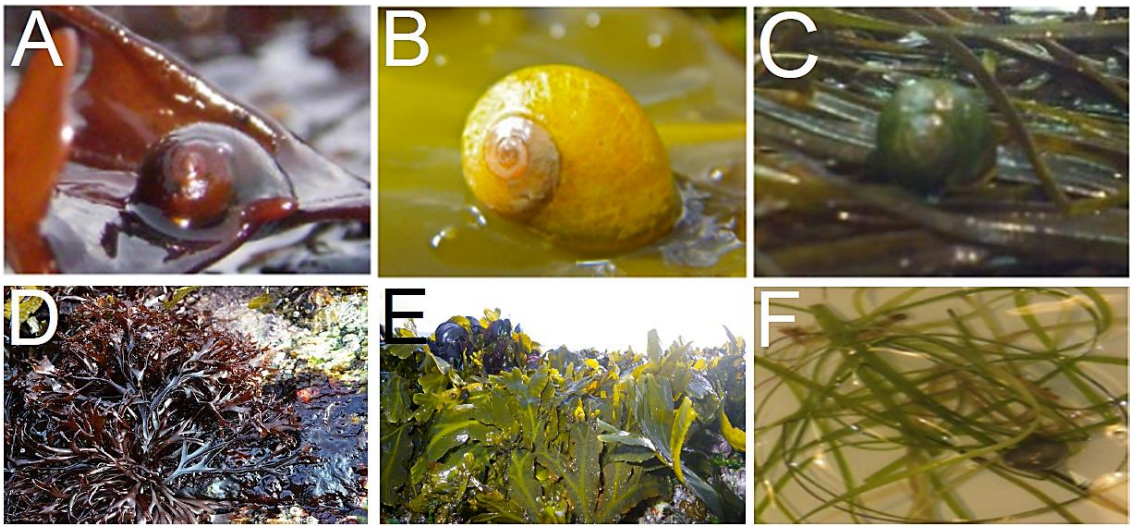

Figure 5. Association between *Littorina fabalis* shell and host algae/seagrass colors, in the Iberian Peninsula. Note the striking resemblance between the red/brownish color of the ME ecotype (A) and the *Mastocarpus* spp. algae (D), the yellow color of the FI ecotype (B) and that of the *Fucus* spp. algae (E), and between the green color of the ZS ecotype (C) and the *Zostera* spp. seagrass (F).

It can also be hypothesized that the smaller size and thinner shell of the ME ecotype results from an adaptation to prevent dislodgement by waves in the exposed shore where it is found, as suggested for the wave-exposed ecotypes of *L.*

saxatilis, which, in addition to its larger foot, presents also a smaller and thinner-shell when compared to the sheltered ecotypes (Butlin et al., 2014). However, the opposite trend is observed in NE, where individuals from moderately exposed shores are larger than those living on more sheltered locations (Tatarenkov & Johannesson, 1999; Johannesson & Mikhailova, 2004). Kemppainen et al. (2005) determined that the increased risk of dislodgment in more exposed habitats creates a selective pressure for a larger size of the LM ecotype in the Swedish shores, because these individuals are able to more effectively withstand crab predation when they fall off their host algae. Although this suggests that selection on size in *L. fabalis* depends on a complex interaction between different factors (e.g. predation, dislodgment risk and protection by host algae).

Concerning the spatial distribution of the ecotypes, in Sweden and Norway, the LM and SS ecotypes are almost parapatric, with a continuous distribution within a range of 150 to 300 meters between the sheltered and moderately exposed extremes, meaning plenty of opportunity for gene flow among ecotypes (Figure 6); whereas in UK, despite their close geographic location (<10 Kilometers (Km)), the ecotypes tend to be allopatrically distributed (Rui Faria, pers. obs.). Although the distribution of the different *L. fabalis* ecotypes in the IP is allopatric (Rui Faria, pers. obs.), they are generally separated by larger distances than in UK (Figure 7). Nevertheless, regardless of their current distribution, current and/or past gene flow between the ecotypes within each country/region is not implausible.

Figure 6. Location of the moderately exposed and sheltered habitats within a single location in Sweden (Lökholmen Island). The LM and SS ecotypes of *Littorina fabalis* present an almost parapatric distribution within a range of 150 to 300 meters between the sheltered and moderately exposed extremes.

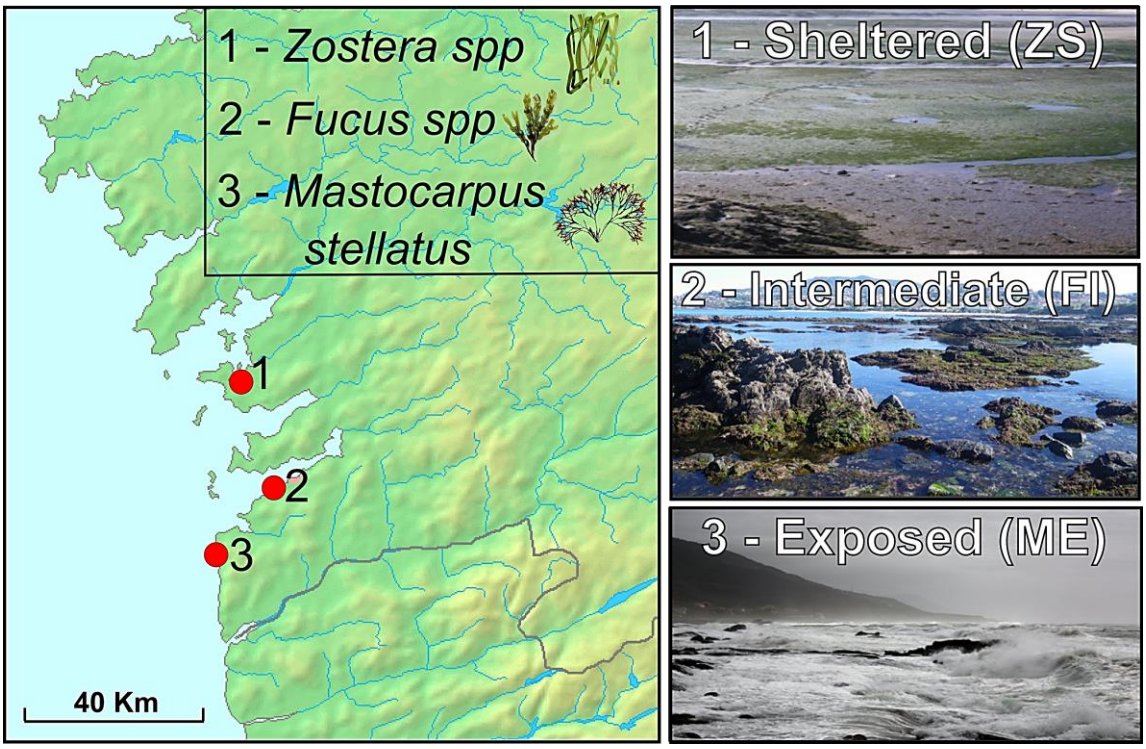

Figure 7. Location of three populations of *Littorina fabalis* ecotypes and respective host algae/seagrass in Galicia (Northern IP) analyzed in this study. The distribution of the ecotypes is not restricted to these sites but we use this example to highlight their allopatric distribution, with populations separated by >25 kilometers. The algae/seagrass, level of wave exposure and ecotype (in brackets) associated with each site is indicated. Images illustrating the habitats of the different sites are also included.

Only a few studies have focused on the genetic differentiation between *L. fabalis* ecotypes and even fewer in *L. obtusata* (but see Schmidt et al., 2007). Evidence for a role of natural selection in the evolution of *L. fabalis* ecotypes comes from the discovery of an association between the different ecotypes and contrasting allelic frequencies at one allozyme locus (Arginine Kinase, ArK) detected in *L. fabalis* populations in two small islands of the Western coast of Sweden (Tatarenkov & Johannesson, 1994), Wales and France (Tatarenkov & Johannesson, 1999), which was recently confirmed by the finding of signatures of a selective sweep in this same gene using sequencing data (Kemppainen et al., 2011). However, the fact that similar signatures of selection in this gene were not found in the IP suggests a different genetic makeup of the *L. fabalis* ecotypes from this region (Tatarenkov & Johannesson, 1999), although this needs to be further studied in more detail because of the inherent problems of allozyme studies.

4. Focal system and main goals

The system comprised by the flat periwinkles and their ecotypes allows the exploration of the mechanisms promoting divergence across the speciation continuum (between different ecotypes and sister species). Although natural selection is most likely playing an important role in ecotype divergence in *L. fabalis*, as well as in reproductive isolation between this species and *L. obtusata*, many of the previous studies were based on a limited number of markers of low resolution, thus lacking the necessary power to adequately tackle this question.

In order to circumvent those limitations, here I will analyze several populations of *L. fabalis* from NE, including the SS and LM ecotypes; and from the IP, including the ME, FI and ZS ecotypes, together with populations of *L. obtusata*, by means of AFLP (Amplified Fragment Length Polymorphism) and microsatellite markers, respectively, complemented by a phenotypic analysis of the shell based on geometric morphometrics. The independent genetic analysis of the two regions (the IP and NE) is justified by the differences in available information from previous studies on both regions (i.e. very limited knowledge for Iberian populations) and the split, evident at mtDNA, between Iberian and Northern European populations (Kemppainen et al., 2009; Faria et al., unpublished results), similarly to *L. saxatilis* (Butlin et al., 2014).

By examining different stages along the continuum of speciation (e.g. ecotypes adapted to different habitats, to different habitats and host algae and sister species), I aim to improve our understanding of the mechanisms responsible for diversification in flat periwinkles, hoping that this knowledge could be applicable to other taxa, allowing to make generalizations about the mechanisms of speciation. To achieve this main goal, three specific objectives were defined:

1. Study the phenotypic differentiation in shell size and shape between *L. fabalis* and *L. obtusata* as well as among the different *L. fabalis* ecotypes, developing a new geometric-morphometrics protocol specific for flat periwinkles.

2. Provide a new battery of polymorphic microsatellite markers specific for flat periwinkles to assess the genetic variation in populations of *L. obtusata* and *L. fabalis* (including ME, FI and ZS ecotypes) from the IP, as well as to detect putative cases of hybridization between *L. obtusata* and *L. fabalis.*

3. Perform an AFLP genome scan between LM and SS ecotypes of *L. fabalis* from NE to evaluate the degree of parallelism of their divergence at different geographic scales (across different countries and across locations within countries) and the proportion of the genome under divergent selection between sheltered and moderately-exposed locations.

Material and methods

1. Prospection of flat periwinkles in the North of Portugal

Despite several descriptions of the distribution of flat periwinkles in NE and in Galicia (Rolán & Templado, 1987; Reid, 1996; Kemppainen et al., 2011), their presence in Portugal, considered the Southern limit of the species, was basically unknown. In order to fill this gap, an initial prospection along the Portuguese coast from Caminha to Nazaré was carried out between 2011 and 2013 (see Supplementary Information).

2. Sampling

A total of 21 sites were selected for sampling, encompassing two main regions: IP and NE. Sampling methodology is described in Supplementary Information.

2.1. Sampling locations in the Iberian Peninsula

In the IP, 918 samples were collected from September 2012 to February 2013 along the Northern coast of Portugal and Galicia (Table 2, Figure 8,). Each location was classified in terms of exposure to wave action, inferred from the presence of different algae. *Mastocarpus stellatus* is typical of more exposed sites of the lower intertidal (locations 1, 2, 11 and 12, Table 2, Figure 8). In the other extreme, the presence of the seagrass *Zostera* spp. is characteristic of very sheltered locations inhabiting sandy/muddy subtracts (locations 6 and 7, Table 2, Figure 8). Meanwhile, the presence of *Fucus* spp. is indicative of intermediate exposure between the previous two (locations 3, 4, 5, 8 and 9, Table 2, Figure 8).

Although Moinhos (location 10, Table 2, Figure 8) is located in a very exposed shore, a barrier of rocks in the lower intertidal allows the existence of *Fucus* spp. in the upper (and protected) part, where *L. obtusata* is usually found, suggesting that the specific place where they have been collected is indeed a sheltered location. As well, although Cabo do Mundo (location 13, Table 2, Figure 8) is also wave-exposed, the specific sampling site is somehow protected by the configuration of the beach and inhabited by *Fucus* spp. However, it is not clear if strong wave action in winter

can pose important selective pressures in this locality, which was classified as unknown in order to be conservative.

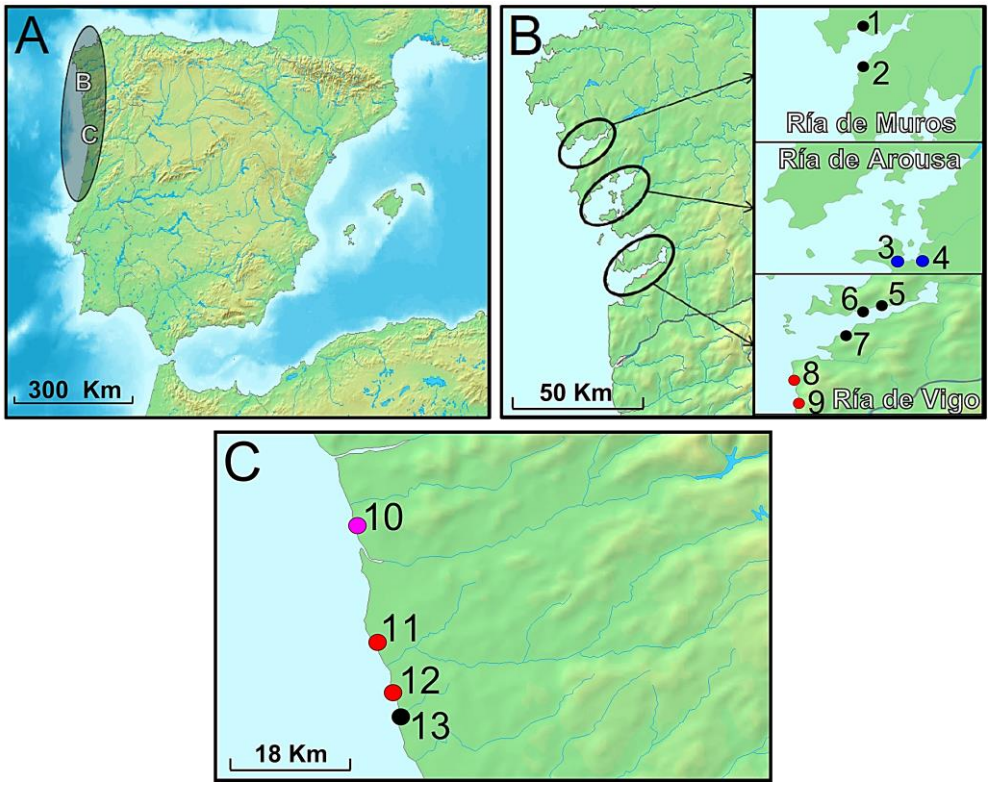

Figure 8. Sampling locations in the IP. (A) Flat periwinkles' distribution is limited to the Northwestern shores of the IP. (B) In Galicia, sampling locations span three main Rías: Muros e Noia, Arousa and Vigo. (C) In Portugal, sampling locations are comprised between South of Viana do Castelo and North of Porto, from top to bottom. This area covers the entire range of habitat conditions associated with *L. fabalis* ecotypes in the region. Point colors indicate the putative sampled ecotype: black for FI, blue for ZS and red for ME; while pink indicates a putative *L. obtusata's* population sampled for comparison purposes. Point numbers correspond to those found in Table 2.

Table 2. Sampling information for the IP. *N* is the number of sampled individuals (918 in total). Location numbers in front of each name follow Figure 8. Putative species and ecotypes were inferred based on the type of habitat where the snails were collected and on their shell appearance (determined on site).

Location	Habitat Type	Sampling Date	*N*	Ecotype	Putative Species
Oia (1)	Exposed	November 2012	24	ME	*L. fabalis*
Silleiro (2)	Exposed	Oct. 2012/Feb. 2013	74	ME	*L. fabalis*
Canido (3)	Intermed. to Exposed	October 2012	93	FI	*L. fabalis*
Cangas (4)	Intermediate	November 2012	119	FI	Mainly *L. fabalis*
Tirán (5)	Intermediate	November 2012	133	FI	*L. fabalis*
Grove 1 (6)	Sheltered	December 2012	60	ZS	*L. fabalis*
Grove 2 (7)	Sheltered	December 2012	23	ZS	Mainly *L. fabalis*
Muros (8)	Intermed. to Exposed	December 2012	25	FI	*L. fabalis*
Abelleira (9)	Intermediate	December 2012	84	FI	Mainly *L. fabalis*
Moinhos (10)	Sheltered	November 2012	85	-	*L. obtusata*
Póvoa (11)	Exposed	November 2012	63	ME	*L. fabalis*
Agudela (12)	Exposed	November 2012	65	ME	*L. fabalis*
Cabo do Mundo (13)	unknown	November 2012	70	FI	*L. fabalis*

2.2. Sampling locations in Northern Europe

In NE, 662 individuals were collected between August and October 2012 in Norway, Sweden and the United Kingdom (UK), from at least two locations within each country (Table 3, Figure 9). The presence of *Ascophyllum* spp. (besides *Fucus* spp.) was used as the criterion to classify these locations as sheltered, while in moderately-exposed locations only *Fucus* spp. was generally found (Table 3), as described in Tatarenkov and Johannesson (1998).

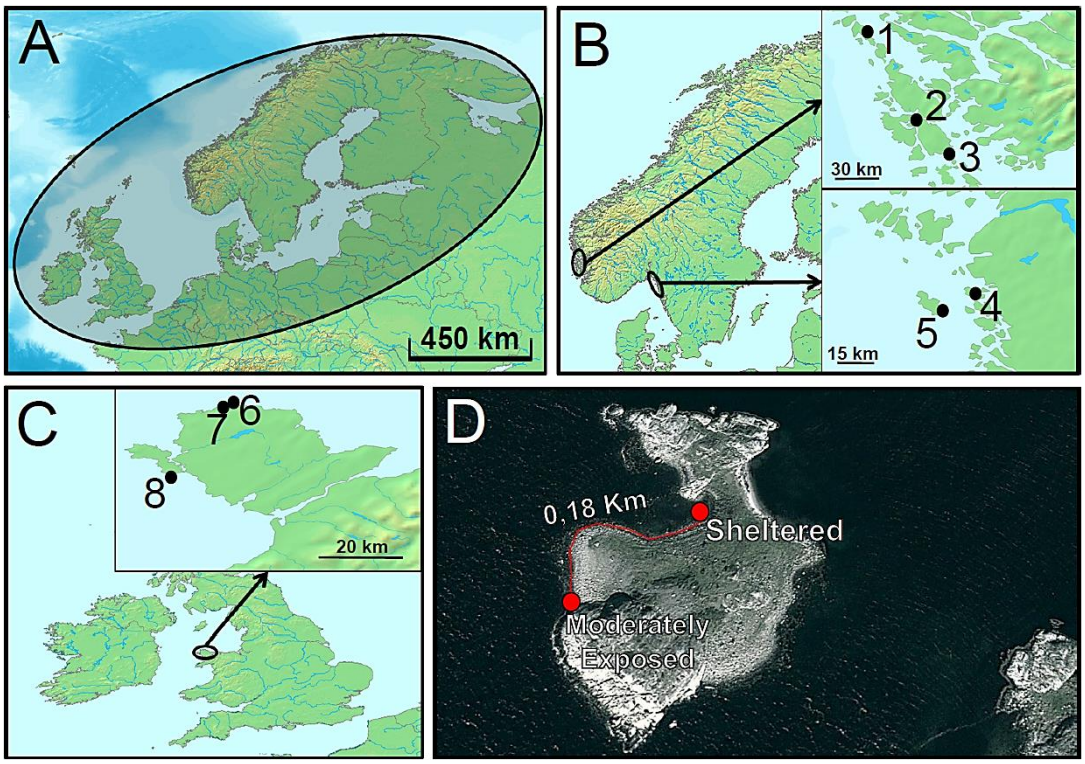

Figure 9. Sampling locations in NE. (A) Flat periwinkles are widely distributed in the North Atlantic, occupying most coastal areas. (B) In Norway, sampling was conducted at: 1 – Sele, 2 – Syltonya and 3 – Hummelsund; and in Sweden at: 4 – Lokholmen and 5 – Ursholmen. (C) In Anglesey (UK), sampling was conducted at: 6 and 7 – Anglesey North and 8 – Anglesey South. Black dots indicate sampling locations, with numbers corresponding to those in Table 3. (D) In Norway and Sweden, moderately-exposed and sheltered sites in each location were separated by less than 1 Km, as exemplified in Lokholmen (point 4, Sweden).

Within each location, a moderately-exposed and a sheltered site were sampled, with two exceptions: i) North Anglesey in UK, where the sheltered (6) and moderately-exposed (7) sites are from different locations (Table 3, Figure 9); and ii) Ursholmen (5) in Sweden where, despite the indication of the existence of one moderately-exposed and one sheltered site (Kerstin Johannesson, personal communication), the high density of *Ascophyllum* spp. observed in both sites rather suggests that they are both sheltered (Table 3, Figure 9). Thus, in Norway and

Sweden, the sampled moderately-exposed and sheltered habitats were geographically very close to each other (<1 Km), while in UK the distance between the two habitats was larger (South Anglesey: 1.5 Km; North Anglesey: 10 Km) (Figure 9). The sampling methodology in NE is also described in Supplementary Information.

Table 3. Sampling summary for NE. Locality numbers in front of each name follow Figure 9. *N* is the number of individuals sampled (662 in total, from three countries). Ecotype was inferred based on the type of habitat where the snails were collected and the abundance of *Ascophyllum* spp. LM, Large-moderately exposed ecotype; SS, Small-Sheltered ecotype

Country	Location	Habitat Type	Code	Sampling Date	*N*	Ecotype
Norway	Sele **(1)**	Moderately-Exposed	Sel_Exp	August 2012	30	LM
Norway	Sele **(1)**	Sheltered	Sel_Shl	August 2012	26	SS
Norway	Syltonya **(2)**	Moderately-Exposed	Syl_ExpA	August 2012	30	LM
Norway	Syltonya **(2)**	Sheltered	Syl_Shl	August 2012	22	SS
Norway	Syltonya **(2)**	Moderately-Exposed	Syl_ExpB	August 2012	34	LM
Norway	Hummelsund **(3)**	Sheltered	Hum_Shl	August 2012	38	SS
Norway	Hummelsund **(3)**	Moderately-Exposed	Hum_Exp	August 2012	33	LM
Sweden	Lokholmen **(4)**	Moderately-Exposed	Lok_Exp	Sept./Oct. 2012	43	LM
Sweden	Lokholmen **(4)**	Sheltered	Lok_Shl	Sept./Oct. 2012	41	SS
Sweden	Ursholmen **(5)**	Moderately-Exposed*	Urs_Exp*	Sept./Oct. 2012	35	LM
Sweden	Ursholmen **(5)**	Sheltered	Urs_Shl	Sept./Oct. 2012	59	SS
UK	Anglesey – North **(6)**	Sheltered	AngN_Shl	September 2012	50	SS
UK	Anglesey – North **(7)**	Moderately-Exposed	AngN_Exp	September 2012	21	LM
UK	Anglesey – North **(7)**	Intermediate	AngN_Int	September 2012	22	SS
UK	Anglesey – North **(7)**	Unknown**	AngN_Unk**	September 2012	50	LM
UK	Anglesey – South **(8)**	Sheltered	AngS_Shl	September 2012	56	SS
UK	Anglesey – South **(8)**	Moderately-Exposed	AngS_Exp	September 2012	72	LM

*Despite previous information on this site as moderately-exposed, the observed high density of *Ascophyllum* spp. rather suggests that it is sheltered. **In this location, the exposure could not be objectively determined. Despite presenting features compatible with a classification between intermediate and moderately-exposed, it was conservatively classified as unknown.

The protocol for sample processing is described in detail in the Supplementary Information. Briefly, individuals were sexed using the dissection microscope (Nikon SMZ1000) and males were classified into *L. fabalis* or *L. obtusata* based on the penis morphology, whereas females were classified based on their shell appearance. Nonetheless, their species status was further evaluated by means of geometric morphometrics and genetic analyses, for which shell and soft tissues were separately preserved.

3. Geometric Morphometrics analysis

In order to identify the differences in shell size and shape between groups of individuals a geometric morphometrics (GM) analysis was performed (Rohlf & Bookstein, 2003) (see Supplementary Information). Shells were positioned following the protocol developed for *L. saxatilis* (Carvajal-Rodríguez et al., 2005) (Figure 10) and photographed with a Nikon SMZ1500 dissection microscope.

Based on a preliminary analysis, different number of landmarks (LM) and semilandmarks (SLM) were used for individuals from the IP and from NE (Figure 10, see Supplementary Information).

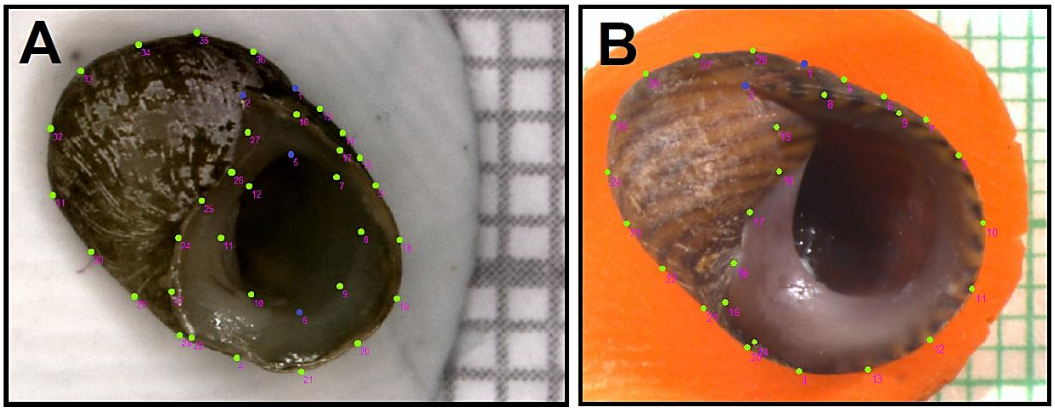

Figure 10. Standard position in which the photographs were taken and placement of the used landmarks. (A) Iberian FI ecotype of *L. fabalis* with 4 LMs (blue dots) and 32 SMLs (green dots). (B) Northern European LM ecotype of *L. fabalis* with 2 LMs (blue dots) and 26 SMLs (green dots). SLMs are equidistantly placed from each other and between two fixed landmarks. Each square of the grid has 1 mm sides.

The software packages tpsUtil v.1.58, tpsDig v.1.40 and tpsRelw v.1.49 (http://life.bio.sunysb.edu/ee/rohlf/software.html) were used to perform the Generalized Procrustes Analysis, based on the superimposition method (Kaliontzopoulou, 2011), following the pipeline described in Figure 11. Size was studied using the centroid size (CS - defined by the squared root of the sum of the square distances of each LM and SLM to the centroid), and shape differences were subdivided into uniform (U1 and U2) and non-uniform (Relative Warps, RWs) components (see Supplementary Information).

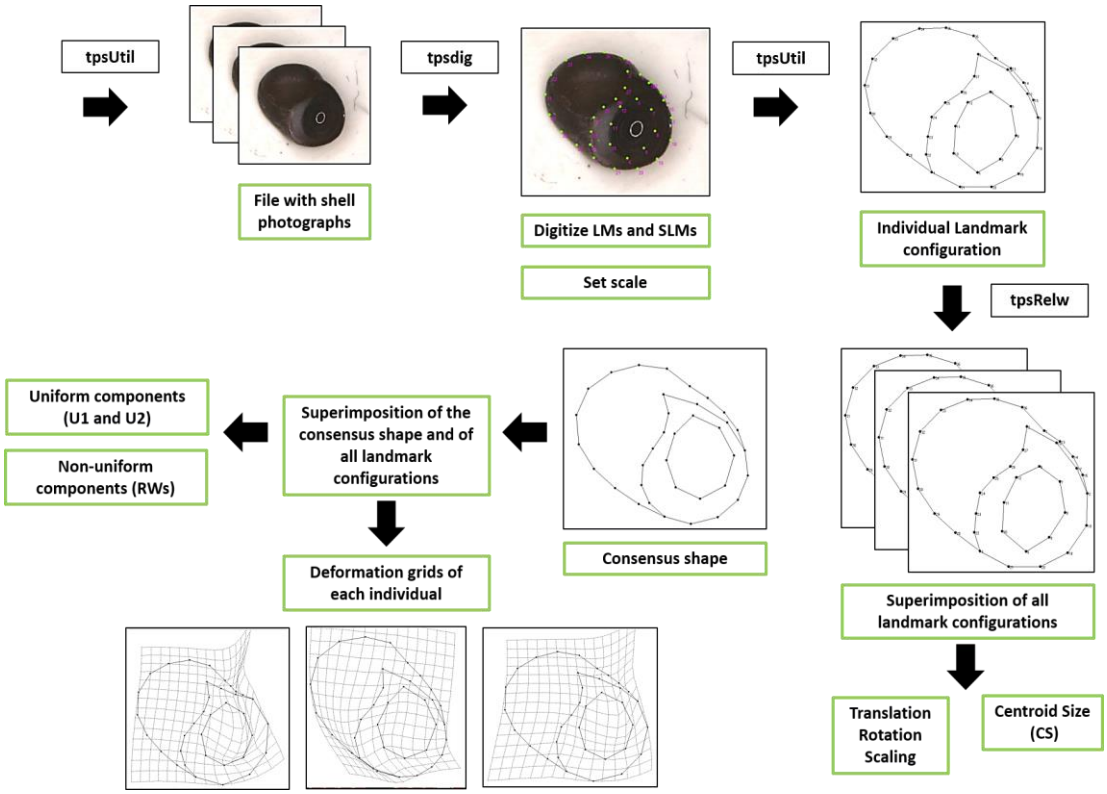

Figure 11. GM analysis pipeline. Illustrative representation of the implemented pipeline with the software used in each step.

In the IP, a total of 184 individuals (only adult males, with the species assigned based on penis morphology) were analyzed (Table 4). Additionally, individuals with shell scars, resulted from crab attacks, were also excluded. For the ME individuals from Silleiro and Oia, it was not possible to remove the soft tissues without damaging the shell. Consequently, they were only included in the genetic analysis.

Table 4. Individuals from the IP included in the GM analysis. Location numbers after each name follow Figure 11. *N* is the number of individuals analyzed for each population. In total: 92 FIs, 32 MEs, 23 ZSs and 37 *L. obtusata.*

Location	*N*	Ecotype	Putative Species
Canido (3)	21	FI	*L. fabalis*
Cangas (4)	15	FI	Mainly *L. fabalis*
Tirán (5)	30	FI	*L. fabalis*
Grove 1 (6)	14	ZS	*L. fabalis*
Grove 2 (7)	9	ZS	Mainly *L. fabalis*
Muros (8)	11	FI	*L. fabalis*
Abelleira (9)	15	FI	Mainly *L. fabalis*
Moinhos (10)	27	-	*L. obtusata*
Póvoa (11)	12	ME	*L. fabalis*
Agudela (12)	20	ME	*L. fabalis*
Cabo do Mundo (13)	10	-	-

In NE, a total of 78 individuals (adults) were analyzed (Table 5). Since sample size in some locations was low, both males and females were included in the analyses.

Table 5. Individuals from NE included in the GM analysis. Locality numbers after each name follow Figure 12. *N*, number of individuals analyzed for each population. In total: 39 SSs and 39 LMs.

Country	Locality	Code	*N*	Ecotype
Norway	Sele (1)	Sel_Exp	16	LM
Norway	Sele (1)	Sel_Shl	8	SS
Sweden	Lokholmen (4)	Lok_Exp	5	LM
Sweden	Lokholmen (4)	Lok_Shl	14	SS
UK	Anglesey – South (8)	AngS_Shl	17	SS
UK	Anglesey – South (8)	AngS_Exp	18	LM

3.1. Data analysis

In the IP, two different analyses, i) including and ii) excluding *L. obtusata* individuals, were performed. To uncover significant differences in the means across ecotypes and across each ecotype and *L. obtusata*, we applied different statistical tests (One-way ANOVA, completed with post-hoc tests and t-tests; see Supplementary Information for details). Additionally, a PCA (Principal Component Analysis) was performed to summarize the different morphological variables (CS, U1, U2, RWs) and determine whether or not the ecotypes (and species) can be accurately distinguished based on shell morphology (see Supplementary Information). The same statistical analyses were performed in the *L. fabalis* ecotypes from NE. All analyses were carried out using STATISTICA v.12 (Sokal & Rohlf, 1994).

4. Genetic analysis

The genetic analysis included two main sections. Firstly, we performed a comprehensive genetic characterization of the *L. fabalis* ecotypes in the IP and investigated the degree of differentiation between this species and *L. obtusata,* identifying putative hybrids, through the development and analysis of highly variable neutral markers (microsatellites). Secondly, benefiting from the more

extensive knowledge concerning the genetic differentiation between the *L. fabalis* ecotypes in NE (e.g. Tatarenkov & Johannesson, 1999; Kemppainen et al., 2009), we performed a genome scan to identify genomic regions underlying adaptive divergence between these ecotypes in different countries by means of AFLP (Amplified Fragment Length Polymorphism) markers.

4.1. DNA extraction

Genomic DNA was extracted from foot tissue using the CTAB-chloroform protocol described in Galindo et al. (2009). DNA quantity and purity were assessed with a Biophotometer (Eppendorf) and adjusted to a final concentration of 20 ng/μL for each individual.

4.2. Microsatellite analysis

In order to investigate the degree of differentiation between the Iberian *L. fabalis* ecotypes and between this species and *L. obtusata,* a battery of highly variable neutral markers (microsatellites) was developed and genotyped for the 13 populations mentioned above (Table 2).

4.2.1. Laboratorial procedures

Microsatellite loci development was performed by GENOSCREEN (Lille, France), following the protocol described by Malausa et al. (2011). Thirty-three primer pairs were selected and initially tested (see Supplementary Information), 17 of which were amplified in three multiplex PCR reactions for 344 individuals (Table 6).

Each individual (20 ng of DNA) was amplified with 4 μL of QIAGEN Multiplex kit, 0.2 μM of each FAM/HEX primer pair and 0.4 μM of each NED primer pair in a final volume of 8 μL. PCR conditions comprised 15 initial minutes at 95°C, followed by 30 cycles of 30 s at 94°C, 90 s at 60°C and 60 s at 72°C, and 30 final minutes at 60°C. One μL of a 1:20 dilution of each PCR product was loaded along with 0.15 μL of GeneScan 400HD ROX size standard (Applied Biosystems) on an ABI 3730 sequencer (Applied Biosystems). Capillary electrophoresis was outsourced to Stab Vida (Setúbal, Portugal).

Table 6. Summary of the 17 microsatellite loci used in this work. Name indicates the name of each locus and they are grouped by multiplex reaction. Size refers to the predicted size in base pairs obtained from the enriched microsatellite libraries (see Supplementary Information). Tm F and Tm R are the melting temperatures of forward and reverse primers, respectively, for which sequences are also indicated.

	Name	Size	Tm F	Tm R	Forward primer (5'-3')	Reverse primer (5'-3')
Multiplex 1						
	PBL8	197	60	62	CCCAGACAATGCAGCCTAC	CGGTAACTGAGTTGTGCAGC
	QVOM	117	62	62	ACATGGGATACGACTACCCG	AGCCTAGCTGCTACGTCCAA
	193Q	215	62	58	TTTGCATACACCCGTCTAACC	GCTATTTCATTAAGCCGCCA
	KJ2E	245	62	60	TCACTTACCTCAAACCTTGCG	CCACAGGCGGGGTGTAAG
	VPVX	198	58	58	CGCTACGCCACTTCGTTTA	AATCGGAGAACAAAACCACG
	881	316	58	62	ACGCCCAGAATTGCCTAAAT	GCTTGTTTATTGACAGGCAGC
Multiplex 2						
	EKYY	145	60	60	TTGTCAAGAATGTTGGTTCCC	ATCCGGAATCGACAAGTGAC
	XENN	242	58	58	CAGCACAAGGCGGTTCAG	TCCTATTTGAAGATGCGGTG
	ZIBW	96	58	58	TTTTGTTAACACGTGGCAGTT	TTGGTGAGTGCGTGCATTAT
	LHYM	192	62	58	TGGTACGGACGAGGCTCTTA	ATTGCTTGAATGCCCGTTAC
	927	241	62	62	CATACAATCCGTCCCTCTCC	TACTCGAACAGGAACGAGGC
	1871	105	60	60	CACCCACCCCTATTACCCA	GGGTTGATGGATGAGTGGAT
Multiplex 3						
	DAEH	242	60	60	ACCGCACAGCTACACGAAG	TCGTGTTTCATGATGCCCTAT
	47	194	62	62	TGTTGCTCTGCAGATTATGACA	GATCGATGCCCTGACATAGC
	EVLS	112	58	62	GTTTTGGTTGAATGTTGGGC	GACAGAAAACAGAAACAACGAAA
	TEM7	237	60	60	CTCATGCTGTTCCTGGTTGA	TGCGTGGTTTAAATTGTTCTTG
	ZR6M	105	60	62	TGAGACATGAAGCCTGTGCT	AATACAATCTGGTGTCTGCGG

Genotyping was manually performed in GeneMapper v.3.7 (Applied Biosystems). It is important to note that, to rule out potential doubts and to confirm genotyping consistency, 321 individuals (out of the total 344) were genotyped twice for all the loci. The *EKKY* locus revealed genotyping ambiguities, due to the amplification of multiple peaks in some populations, and it was consequently removed from further analyses. The *DAEH* locus failed to amplify in *L. obtusata*, but it was maintained in the analysis of genetic variability/differentiation within *L. fabalis*. In addition, because of the difficulties in objectively genotype the *193Q* locus in Agudela, Oia, Silleiro and Grove 2, all these individuals were coded as missing data for this locus.

4.2.2. Data analysis

Hardy-Weinberg equilibrium (HWE) for each population/locus pair and linkage disequilibrium (LD) between locus pairs for all populations were evaluated through exact probability tests in GENEPOP v.4.2 (Rousset, 2008), using a Markov Chain (MC) algorithm with default parameters. A Bonferroni correction (Rice, 1989) was subsequently applied to account for multiple tests. Significant Hardy-

Weinberg disequilibria were further inspected to distinguish between possible genotyping errors (null alleles, stuttering and large allele dropout) using MICRO-CHECKER v.2.2.3 (van Oosterhout et al., 2004).

Genetic diversity was evaluated through several parameters. Average expected (*He*), observed (*Hobs*) and non-biased heterozygosity (*Hnb*); percentage of polymorphic loci, either taking 1% (P_{99}) or 5% (P_{95}) as the minimum allelic frequency to consider an allele as a true polymorphism rather than an artifact; and mean number of alleles per locus (*A*) for each population were estimated using GENETIX v.4.05 (Belkhir et al., 1996). Mean allelic richness (*Ar*) and private allelic richness (*PAr*) were estimated using the rarefaction method implemented in HP-Rare v.1.1 (Kalinowski, 2005). Since *PAr* of a given population does not only depends on its own genetic variability but also on the diversity of the other populations in the dataset, and since the number of populations analyzed for *L. fabalis* and *L. obtusata* differs considerably (two vs. 11), *PAr* was separately calculated for each species.

Population structure was investigated using the Bayesian clustering method implemented in STRUCTURE v.2.3.4 (Falush et al., 2007). Two separate analyses were performed. First, to inspect the differences between *L. fabalis* and *L. obtusata*, as well as the presence of putative hybrids, information from all genotyped individuals was included for 14 loci (besides the *EKKY*, *193Q* and *DAEH* were removed to avoid distorted estimates of hybridization/introgression due to null alleles), with the number of clusters (*K*) ranging from 1 to 13 (the total number of locations sampled). Second, to assess the genetic substructure within *L. fabalis*, STRUCTURE was run including only the individuals we were certain of being true *L. fabalis* (using penis morphology – males, and information from the previous STRUCTURE run - males and females). In this case, the genotypes for 15 loci were included (besides the *EKKY*, the *193Q* was removed due to the existence of null alleles in *L. fabalis*), and 1 to 11 (the total number of *L. fabalis* sampled locations) clusters (*K*) were considered. For the two analyses, ten replicate runs were performed for each *K*, with 10,000,000 iterations (100,000 as burn-in), assuming

an admixture model, correlated allele frequencies and without population prior information.

The method of Evanno et al. (2005), implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012), was then employed to determine the *K* that best fitted the data. The results from the multiple replicates of the best *K* value were combined using the Greedy algorithm in CLUMPP v.1.1.2 (Jakobsson & Rosenberg, 2007) and the obtained output was plotted using DISTRUCT v.1.1 (Rosenberg, 2004). For the *L. fabalis* dataset, we also used an empirical approach as suggested in the STRUCTURE manual (http://pritch.bsd.uchicago.edu/structure.html), which defines the best *K* as the highest among those with a similarly high posterior probability, in which at least one individual is strongly assigned to each cluster (*Q*>80).

Differentiation between populations was also assessed by means of F_{ST} (Weir & Cockerham, 1984) and R_{ST} (Slatkin, 1995) between all population pairs using FSTAT v.2.9.3.2 (Goudet, 1995) and GENEPOP, respectively. The correlation between pairwise F_{ST} and R_{ST} was tested by means of a Spearman's Rank Correlation Coefficient (Sokal & Rohlf, 1994). Average differentiation between species and ecotypes was estimated as the mean of all pairwise values including populations from the two species or from each pair of ecotypes, respectively.

The correlation between genetic and geographic distance, i.e. isolation by distance (IBD), among *L. fabalis* populations was tested by means of a Mantel test (Mantel, 1967), and its significance obtained with 10,000 permutations, using GENEPOP. Both F_{ST} as well as transformed values of differentiation using Slatkin's (1995) linearized F_{ST} ($F_{ST}/(1-F_{ST})$) were used. Geographic distances between sampling locations were calculated as the shortest distance along the coast according to Google Maps (https://maps.google.com/), with both linear and log transformation of the geographic distances tested against genetic distances. A second analysis of IBD in *L. fabalis* was performed after excluding ME populations, which seem to be affected by stronger drift than the populations from the remaining ecotypes (see Discussion).

Neighbor-joining (NJ, Saitou & Nei, 1987) trees (population- and individual-based) were constructed based on Nei's D_A distance (Nei et al., 1983) in POPULATIONS v.1.2.31 (http://www.cnrs-gif.fr/pge/bioinfo/liste/index.php?lange=fr), with node support estimated through 1000 bootstrap replicates (over loci). FIGTREE v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the trees.

4.3. AFLP analysis

The detection of loci under selection (i.e. outlier loci) by means of AFLP genome scans is a widespread methodology in studies of adaptation and speciation (Nosil et al., 2009a; Butlin, 2010) and it has been successfully applied to different populations within the genus *Littorina* (Wilding et al., 2001; Galindo et al., 2009, 2013; Butlin et al., 2014). Loci are classified as "outliers" when they exhibit significantly greater genetic differentiation (i.e. F_{ST}) than neutral expectations (obtained through simulations); otherwise they are classified as "nonoutliers". Here, we performed a genome scan using AFLPs to investigate the level of divergence between *L. fabalis* ecotypes in NE (Norway, Sweden and the UK) and to identify putative loci underlying ecotype divergence between sheltered and moderately-exposed habitats, as well as the degree of parallelism in such divergence (i.e. proportion of outlier loci shared) at different scales (within country and among countries).

4.3.1. Laboratorial procedures

The general AFLP protocol comprises four main steps: i) digestion with two restriction enzymes (a 4 bp and a 6 bp cutter), ii) ligation with double-stranded adapters complementary to the restriction enzymes' recognition sites, iii) pre-selective PCR with primers containing one selective nucleotide on the 3' end, and iv) selective PCR with primers containing three selective nucleotides (Vos et al., 1995). Here, we applied the specific protocol developed for *L. saxatilis* by Butlin et al. (2014), with minor modifications (see Supplementary Information), among which, I would like to highlight the use of EcoRI (6 bp) and MseI (4 bp) restriction enzymes because according to other L. saxatilis studies (Galindo et al., 2009;

Galindo et al., 2013), they allow more loci to be genotyped when compared to the combination used by Butlin et al. (2014) (EcoRI (6 bp) and PstI (6 bp).

Four selective PCRs (Eco+ACT/Mse+CAA; Eco+AAG/Mse+CAA; Eco+ACT/Mse+CAC; Eco+AAG/Mse+CAC; see Supplementary Information) were performed in a total of 379 individuals from seven localities across three countries. For each selective PCR, 0.8 μL were analyzed on an ABI 3130 sequencer (Applied Biosystems) along with 0.2 μL of GeneScan 500ROX size standard (Applied Biosystems). Electropherograms were analyzed with GeneMapper v.3.7. Loci were manually assigned by defining bins (fragment-length classes) from the overlapping electropherograms of all the samples. Bins were created between 75 and 500 bp and only peaks >50 rfu (relative fluorescent units) were considered. For each sample, fluorescence intensity of the peaks (peak height) within each bin was also determined. This step was repeated for each of the four primer combinations (selective PCRs). The R-script AFLPSCORE (Whitlock et al., 2008) was used to transform peak heights into binary (0/1) genotype data based on quality thresholds (locus selection and phenotype-calling thresholds) determined from the data of replicated samples. AFLPSCORE was also used to estimate the mismatch error rate by comparing the dissimilarity between sample replicates for each combination (Whitlock et al., 2008).

4.3.2. Data analysis

Based on all genotyped AFLP loci, heterozygosity and percentage of polymorphic loci, taking 5% (P_{95}) as the minimum allelic frequency to consider an allele as a true polymorphism, were calculated using AFLP-SURV v.1.0 (Vekemans et al., 2002). The same software was used to calculate genetic differentiation (F_{ST}) and Nei's genetic distances (D), following Lynch & Milligan (1994), using a Bayesian method that assumes a non-uniform prior distribution of allele frequencies (Zhivotovsky, 1999).

The detection of loci under selection (outlier loci) between moderately-exposed and sheltered sites was then performed in an independent manner within each

locality: Hum_Exp/Hum_Shl; Sel_Exp/Sel_Shl; Syl_ExpA/Syl_Shl; Lok_Exp/Lok_Shl; AngN_Exp/AngN_Shl; AngS_Exp/AngS_Shl; as well as Urs_Exp*/Urs_Shl, despite the doubts concerning the exposure in Urs_Exp (Table 3). In Syltonya (Norway), because two exposed locations were sampled, we selected the most exposed one (Syl_ExpA) for outlier detection.

The outlier detection was performed applying the two most commonly used methods in the literature (Pérez-Figueroa et al., 2010): BAYESCAN v.2.0 (Foll & Gaggiotti, 2008) and MCHEZA (Antao & Beaumont, 2011), more and less stringent, respectively.

BAYESCAN first calculates population-specific and locus-specific F_{ST}, and then estimates the posterior probabilities of two alternative models (including or excluding the effect of selection) for each locus using a reversible-jump Markov Chain Monte Carlo (MCMC) approach. Ten pilot runs (10,000 iterations) were performed to tune the model parameters, followed by 400,000 iterations (100,000 as burn-in, 20 as thinning interval and 20,000 as sample size). Loci were identified as outliers when the posterior probability was greater than 0.97, but a correction for multiple tests (false discovery rate - FDR; Benjamini & Hochberg, 1995) was applied to avoid overestimating the proportion of loci that are under divergent selection.

MCHEZA is adapted from the DFDIST program (http://www.maths.bris.ac.uk/~mamab/stuff/), which is based on the method developed by Beaumont and Nichols (1996). The program generates loci obtained through coalescent simulations using a neutral model with two symmetrical islands. Then, the distribution (F_{ST} conditional on heterozygosity) of simulated loci is compared to the empirical data and loci with F_{ST} significantly greater (p<0.05) than the simulated F_{ST} are classified as outliers. The main advantage of MCHEZA compared to DFDIST is that it allows the estimation of the mean neutral F_{ST} while taking into account loci that might be under selection. MCHEZA also introduces support for multi-test correction (FDR method) to reduce the number of false positives. For each locality, 200,000 simulations were performed, with a theta of

0.04 (theta= 2*2*Ne*mu, see DFDIST manual). Two different datasets were created after the detection of outliers with MCHEZA: one with "outliers" and another with "nonoutliers".

AFLPDAT (Ehrich, 2006) was used to create the input files for STRUCTURE, which was also independently run for "outliers" and "nonoutliers" to determine the population structure. Five replicate runs of 500,000 iterations (100,000 as burn-in), for each *K* (from 1 to 15) were performed, assuming an admixture model, correlated allele frequencies and without population prior information.

As for the microsatellite dataset, the method developed by Evanno et al. (2005) implemented in STRUCTURE HARVESTER was employed to determine the best *K*. Additionally, STRUCTURE runs were separately performed for each region (Norway, Sweden and UK) with the "outliers" and "nonoutliers" dataset, using the same conditions as before, but varying the *K* value according to the number of sampled populations within each country (from 1 to 7 in Norway, from 1 to 4 in Sweden, and from 1 to 6 in UK). Results from the multiple replicates of the best *K* were combined with CLUMPP and outputs plotted with DISTRUCT, in the same manner as for the microsatellite dataset.

NJ trees for the "outliers" and "nonoutliers" datasets were constructed based on Nei's *D* distance using the NEIGHBOR routine implemented in the PHYLIP package (Felsenstein, 1981). Ten thousand bootstraps (over loci) were performed using AFLP-SURV. The CONSENSE routine in PHYLIP was used to determine the bootstrap percentage supporting each branch of the tree. Trees were visualized using FIGTREE.

Results

1. Species and ecotype composition of the sampling locations in the Iberian Peninsula

The morphological analysis of the penis largely confirmed the initial assessment of species based on the visual inspection of shell shape and size made in the field, thus revealing a high degree of concordance between this preliminary classification and the one based on penis morphology. The only exception occurred in Cabo do Mundo (Portugal) where individuals had been initially classified as *L. fabalis* but the subsequent inspection of the penis morphology revealed that the majority of the individuals were *L. obtusata* (Table 7). Although in most locations solely males from one of the species were sampled, the two species were found together in Cabo do Mundo (8% of *L. fabalis*), Abelleira (90 of *L. fabalis*), and Grove 2 and Cangas (91% of *L. fabalis*) (Table 7, Figure 12).

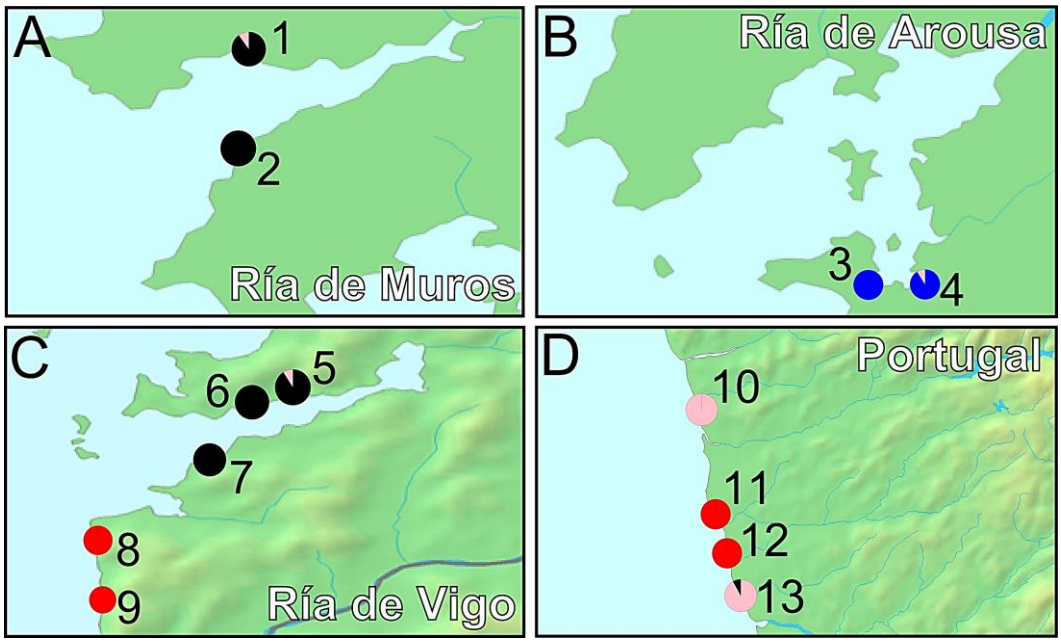

Figure 12. Representation of the frequencies of males from *L. fabalis* and *L. obtusata* sampled at each location. The 13 locations correspond to the ones described in the Methods and Table 7, sampled across three Galician rías (A, B and C) and Northern Portugal (D). Pink represents *L. obtusata*, red stands for *L. fabalis* ME ecotype, blue for *L. fabalis* ZS ecotype and black for *L. fabalis* FI ecotype.

Table 7. Characterization of the sampled locations in the IP in terms of sex and species composition. *N,* represents the total number of individuals collected. Species was determined according to the penis morphology. Note that Cabo do Mundo was initially classified in the field as *L. fabalis* but the analysis of the penis morphology revealed a typical *L. obtusata's* penis.

Location	*N*	% of males	Percentage of *L. fabalis* males
Abelleira (1)	84	45 %	90 %
Muros (2)	25	48 %	100 %
Grove 1 (3)	60	52 %	100 %
Grove 2 (4)	23	48 %	91 %
Tirán (5)	133	46 %	100 %
Cangas (6)	119	48 %	91 %
Canido (7)	93	53 %	100 %
Silleiro (8)	74	43 %	100 %
Oia (9)	24	50 %	100 %
Moinhos (10)	85	49 %	0 %
Póvoa (11)	63	37 %	100 %
Agudela (12)	65	43 %	100 %
Cabo do Mundo (13)	70	36 %	8 %

2. Geometric Morphometrics analysis

2.1. Flat periwinkles from the Iberian Peninsula

The analysis revealed that the first three relative warps (RW1, RW2 and RW3) explain 76% of the variation (RW1=55.64%; RW2=11.58% and RW3=9.59%) within the Iberian populations. Thus, in subsequent statistical analyses only these three RWs were included, along with centroid size (CS) and the uniform components (U1 and U2). The Shapiro-Wilk test performed for each group (*L. obtusata* and three *L. fabalis* ecotypes) showed that all variables (CS, U1, U2 and RW1-3) are normally distributed (p>0.05).

A clear separation between *L. fabalis* and *L. obtusata* is revealed when plotting CS against RW1 (Figure 13A). The separation between the two species is mainly explained by differences in CS, with *L. fabalis* individuals generally presenting the smallest size (mean=1.6123 ± 0.2500) and *L. obtusata* the largest (mean=2.8092 ± 0.2592) (Figure S2). Individuals from Cabo do Mundo appear in an intermediate position between the two species (mean=2.0791 ± 0.1907) (Figure S2). Differences are significant between the three categories (*L. fabalis*, *L. obtusata* and Cabo do

Mundo) (*t*-test; $p<0.0001$). Despite that when plotting U1 against U2 there is certain overlap between the species (Figure 13B), Cabo do Mundo also appears in an intermediate position. Since penis morphology is the most informative trait for classifying males into one of the species, in the following statistical analyses the population of Cabo do Mundo was included in the *L. obtusata* group.

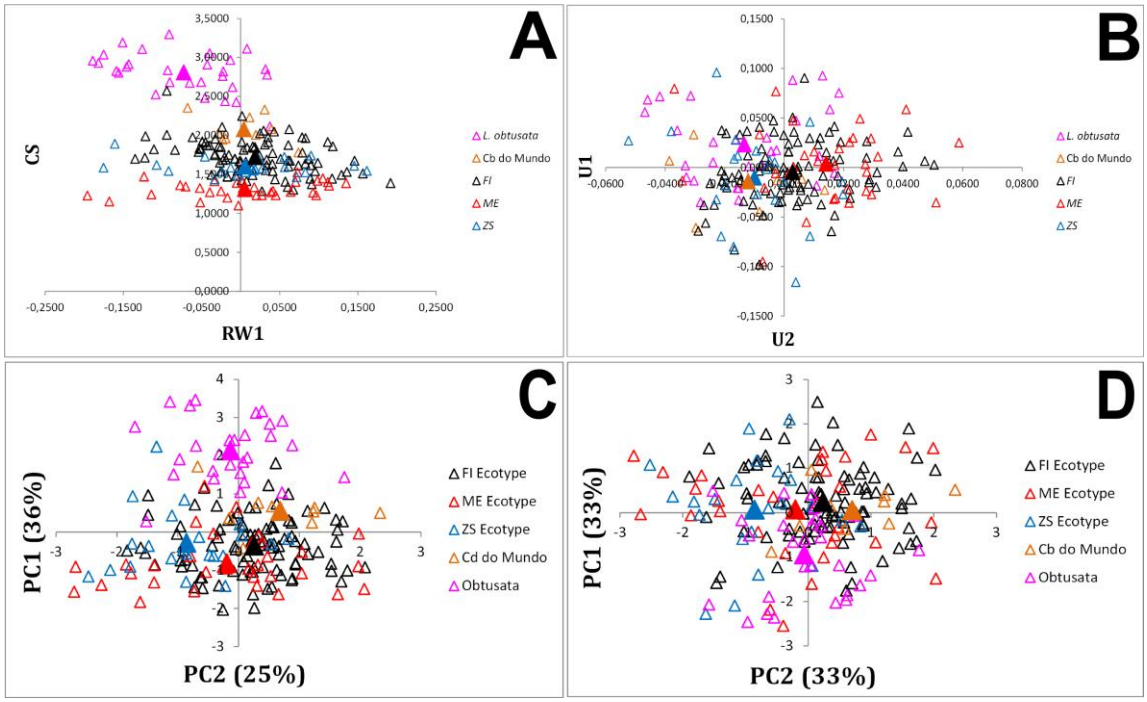

Figure 23. Geometric-morphometrics analysis of Iberian populations. A) Plot of CS against RW1 for *L. fabalis* ecotypes and *L. obtusata*, with Cabo do Mundo as a different group. B) Plot of U1 against U2. C) Plot of PC1 against PC2. D) Plot of PC1 against PC2 excluding CS as a variable. In all the panels, the average position of each group is represented by a larger and color-filled triangle.

A one-way ANOVA shows that *L. obtusata* and the three *L. fabalis* ecotypes are significantly different from each other for all the tested variables (CS, U1, U2 and RW1-3) (Table S4). Post-hoc (Tukey HSD) pairwise tests between all groups (Table 8) revealed a significant difference between the CS of *L. obtusata* and all *L. fabalis* ecotypes, and also between the different *L. fabalis* ecotypes: FI presents the largest size (CS=1.7207 ± 0.2248), followed by ZS (Mean CS=1.5969 ± 0.1024) (not significantly different from FI), and ME (Mean CS=1.3115 ± 0.0999) (the smallest, significantly different from both ZS and FI) (Figure S2). In terms of CS, the difference between populations of the same ecotype is generally smaller than between populations of different ecotypes (Table S5); except for a few FI populations that present a CS more similar to ZS than to other FI populations (see Figure 13A).

Table 8. Tukey HSD tests between each *L. fabalis* ecotype (FI, ME, ZS) and *L. obtusata* (Obt) for all analyzed variables. Numbers between brackets indicate the number of analyzed individuals. Values in bold at the diagonal indicate the mean value for the corresponding group.

	CS				RW1			
Ecotype	FI (92)	ME (32)	ZS (23)	Obt (37)	FI (92)	ME (32)	ZS (23)	Obt (37)
FI	**1.7207**				**0.0184**			
ME	0.0000*	**1.3120**			0.8300	**0.0057**		
ZS	0.1300	0.0000**	**1.5969**		0.8700	1.0000	**0.0053**	
L. obtusata	0.0000**	0.0000**	0.0000**	**2.6316**	0.0000**	0.0400*	0.0100*	**-0.0540**

	RW2				RW3			
Ecotype	FI (92)	ME (32)	ZS (23)	Obt (37)	FI (92)	ME (32)	ZS (23)	Obt (37)
FI	**0.0082**				**0.0080**			
ME	0.1100	**-0.0070**			0.9600	**0.0050**		
ZS	0.0000**	0.5500	**-0.0299**		0.0100*	0.1300	**-0.0131**	
L. obtusata	0.9200	0.4800	0.0100*	**0.0043**	0.0000**	0.0200*	0.9800	**-0.0161**

	U1				U2			
Ecotype	FI (92)	ME (32)	ZS (23)	Obt (37)	FI (92)	ME (32)	ZS (23)	Obt (37)
FI	**-0.0045**				**0.0029**			
ME	0.6900	**0.0037**			0.0100*	**0.0141**		
ZS	0.9000	0.5500	**-0.0102**		0.0100*	0.0000**	**-0.0100**	
L. obtusata	0.0400*	0.6200	0.0500	**0.0143**	0.0000**	0.0000**	0.9200	**-0.0131**

*Significant values (p<0.05); ** Highly significant values (p<0.001)

In terms of shape, U1 reveals a significant differentiation between the FI ecotype and *L. obtusata*, while U2 is significantly different between all analyzed groups except between the ZS ecotype and *L. obtusata*. Concerning the non-uniform components of shape, RW1 is significantly different between *L. obtusata* and all *L. fabalis* ecotypes but not among these. RW2 shows significant differences between ZS and all the other groups except ME; whereas RW3 is significantly different between *L. obtusata* and the *L. fabalis* ecotypes except ZS, and between only two of the *L. fabalis* ecotypes: ZS and FI.

After the exclusion of *L. obtusata* individuals, the three *L. fabalis* ecotypes remained significantly different from each other regarding CS, U2, RW1 and RW3. The observed trends are similar to the previous analysis, both in terms of size (ME is the smallest ecotype: CS=1.3234 ± 0.1211, followed by ZS: 1.5858 ± 0.1175, and FI: 1.7210 ± 0.2260; Tukey HSD post-hoc tests, FI-ME: p=0.000022, FI-ZS: p=0.014092, ME-ZS: p=0.000022); and shape (except for RW2, for which the differences between ME and ZS are now significant; p=0.018) (Table S6).

The first two components of the PCA explained 61% of the variation (PC1=36% and PC2=25%; Figure 13C), although only the first one was significant. When CS was removed from the analysis (PC1=33% and PC2=33%; Figure 13D) none of them

was significant. The plot of these two components confirmed the separation between the species for PC1 when CS was included in the analysis. The individuals from Cabo do Mundo are closer to *L. obtusata* than to the three *L. fabalis* ecotypes (Figure 13C). However, the separation between the groups is less clear when only shape is considered (Figure 13D), although *L. obtusata* remains the most differentiated group, while Cabo do Mundo individuals appear more intermixed within the *L. fabalis* ecotypes.

2.2. Littorina fabalis populations from Northern Europe

The analysis of the North European populations of *L. fabalis* revealed that the first three relative warps explain 75 % of the observed variation (RW1=38.40%; RW2=23.24% and RW3=13.29%), and so only these three RWs, together with U1, U2 and CS were included in subsequent statistical analyses.

The Shapiro-Wilk normality tests showed that these six different variables also conformed to a normal distribution in NE *L. fabalis* populations (p>0.05). Since, in a factorial ANOVA, no significant differences were observed between males and females from this region (*F-value*=1.321; p=0.271; Table S7), both sexes were pooled in subsequent analyses.

A clear separation between the LM and SS ecotypes was obtained by plotting CS against RW1 (Figure 14A). Centroid size is significantly different between the two ecotypes (Tukey HSD pairwise tests, p=0.000115; Table 9) with LM always presenting a larger mean CS than SS (2.6470 ± 0.2182 vs. 1.9500 ± 0.2557). Differentiation between the two ecotypes is also clear when plotting U1 against U2 (Figure 14B), with the main separation observed at the U1, as revealed by the significant differences between the two ecotypes for all variables except U2 (Table 9).

Table 9. Tukey HSD tests between *L. fabalis* ecotypes in NE. LM ecotype (exposed) and SS ecotype (sheltered). Numbers between brackets indicate the number of analyzed individuals. Values in bold indicate the average for each ecotype.

	CS		**RW1**	
Ecotype	Sheltered (39)	Exposed (39)	Sheltered (39)	Exposed (39)
Sheltered	**1.9500**		**-0.0338**	
Exposed	0.0001**	**2.6470**	0.0001**	**0.0338**
	RW2		**RW3**	
Ecotype	Sheltered (39)	Exposed (39)	Sheltered (39)	Exposed (39)
Sheltered	**-0.0005**		**0.0163**	
Exposed	0.9252	**0.0005**	0.0001**	**-0.0163**
	U1		**U2**	
Ecotype	Sheltered (39)	Exposed (39)	Sheltered (39)	Exposed (39)
Sheltered	**-0.0204**		**-0.0029**	
Exposed	0.0001**	**0.0204**	0.1567	**0.0029**

*Significant values (p<0.05); ** Highly significant values (p<0.001)

Despite significant differences in CS between locations within each ecotype (with the exception of the differences between Norway and Sweden for the LM ecotype, t=1.27755, p=0.216799, Table S8), the CS of the SS populations is always more similar to the CS of other SS populations than to the CS of the LM populations from any site and vice-versa, even if in some comparisons they inhabit regions that are about 1000 Km apart.

The first two components of the PCA explain 66 % of the variation, despite only the first one being significant (PC1=41% and PC2=25%; Figure 14C), whereas none was significant when the CS was removed from the analysis (PC1=33% and PC2=33%; Figure 14D). In both cases, the two ecotypes are clearly differentiated at PC1, suggesting that besides the marked differences in size, the two ecotypes present substantial differences in shape that allow their accurate discrimination through PCA.

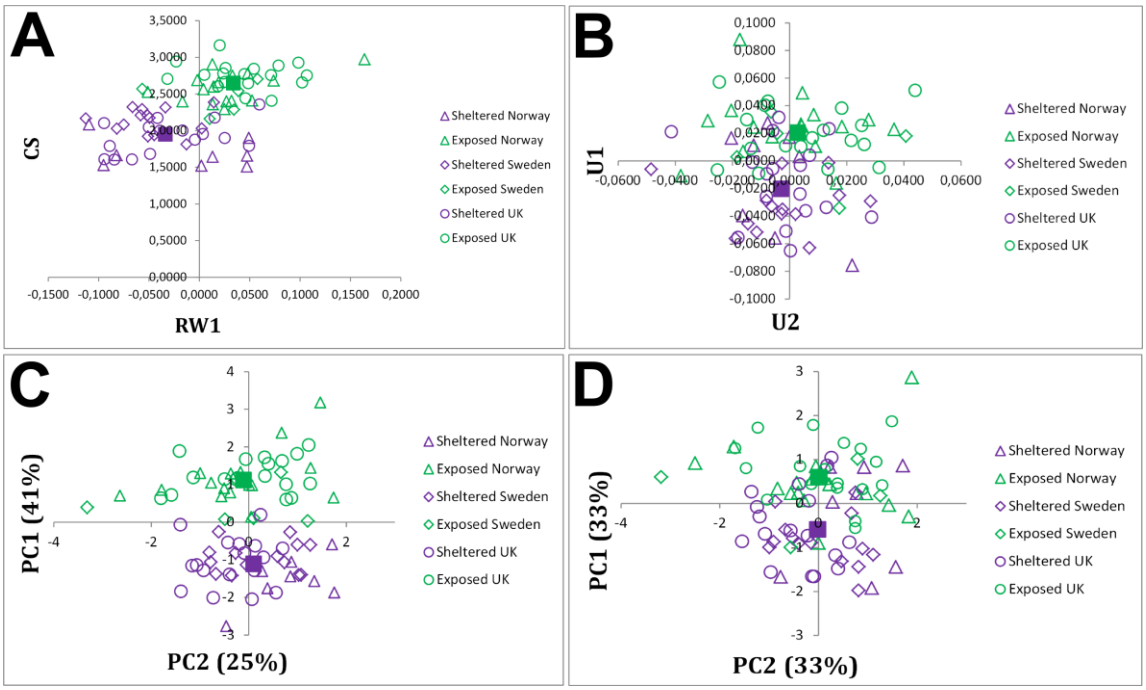

Figure 14. Geometric-morphometrics analysis of *L. fabalis* ecotypes from NE. A) Plot of CS against RW1 for *L. fabalis* ecotypes and *L. obtusata*, with Cabo do Mundo as a different group. B) Plot of U1 against U2. C) Plot of PC1 against PC2. D) Plot of PC1 against PC2 excluding CS as a variable. In all the panels, the average position of each ecotype is represented by a large color-filled square.

3. Genetic characterization of flat periwinkles

3.1. Genetic analysis of flat periwinkles from the Iberian Peninsula using microsatellite loci

3.1.1. Genetic diversity

Thirty seven out of 176 HWE tests for each locus-population combination revealed significant deviations but, after the Bonferroni correction, only five remained significant (p<0.0002) (Abelleira: *QVOM*, F_{IS}=0.4877; Cabo do Mundo: *193Q*, F_{IS}=0.7979; *XENN*, F_{IS} =0.5782; 47, F_{IS}=0.5529; and Grove 2: *EVLS*, F_{IS}=0.7863). However, since HWE (as well as LD) can be greatly affected by the "Wahlund effect" (Sinnock, 1975) (i.e. a reduction in heterozygosity caused by population structure within a sampling site), this analysis was repeated according to the following guidelines: i) after removing the males that presented a penis morphology more compatible with their classification as *L. obtusata* (one from Grove 2 and one from Cangas) in *L. fabalis* populations; ii) after removing the males that presented a penis morphology more compatible with their classification as *L. fabalis* (two

individuals) in Cabo do Mundo (composed mostly *L. obtusata*); and iii) populations where females presented less than 90% of membership (based on the STRUCTURE analysis) to the main cluster formed by the males of the same location (one from Abelleira, one from Grove 2, one from Cangas, two from Oia and four from Cabo do Mundo). After this procedure, we identified 27 significant HWE tests out of 173 (Table S9), with only two remaining significant after the Bonferroni correction (p<0.0002) (Abelleira: *QVOM*, F_{IS}=0.4482; and Grove 2: *EVLS*, F_{IS}=0.7797), possibly due to existence of null alleles, as suggested by MICRO-CHECKER.

LD tests for each locus pair across all populations showed significant loci associations in 18 out of 120 tests, but only three remained significant after the Bonferroni correction (*QVOM-PBL8*, *ZIBW-1871*, and *XENN-ZR6M*; p<0.0004). After the correcting for the Wahlund effect, only seven out of 121 tests showed significant associations, with none remaining significant after the Bonferroni correction (Table S10).

Average heterozygosity (*He*, *Hnb* and *Hobs*), number of alleles per locus (*A*), allelic richness (*Ar*) and private allelic richness (*PAr*) are higher in *L. fabalis* when compared with *L. obtusata*, while the opposite is observed for the percentage of polymorphic loci (Table 10).

Table 10. Genetic diversity in flat periwinkles of the IP. *He*, expected heterozygosity; *Hnb*, non-biased heterozygosity; *Hobs*, observed heterozygosity; *A*, number of alleles per locus; *Ar*, allelic richness; *PAr*, private allelic richness; P_{95}, percentage of polymorphic loci using the 95% criterion; P_{99}, percentage of polymorphic loci using the 99% criterion. Average values across all the populations from one species and for each *L. fabalis* ecotype are presented.

		He	***Hnb***	***Hobs***	***P*95**	***P*99**	***A***	***Ar***	***PAr***
L. fabalis		0.4710	0.4812	0.4456	0.8027	0.8545	4.7852	4.5473	0.1845
	FI	0.5166	0.5277	0.4797	0.8375	0.9000	5.4750	5.0400	0.3040
	ME	0.4040	0.4125	0.3884	0.7407	0.8052	3.7979	3.7600	0.0950
	ZS	0.4909	0.5027	0.4750	0.8396	0.8396	5.0354	4.8900	0.0650
L. obtusata		0.4603	0.4688	0.4307	0.8667	0.9000	4.0000	3.7100	1.5300

When we focus on the three *L. fabalis* ecotypes described for the IP, FI shows the highest genetic diversity (except for P_{95}) and ME the lowest, except for *PAr* that is slightly higher in ME (*PAr*=0.0950) than in ZS (*PAr*=0.0650) (Table 10).

At the population level, a North to South gradient is observed regarding heterozygosity in *L. fabalis*, with FI samples from the Ría de Muros e Noia (Muros and Abelleira) showing the highest heterozygosity; whereas, at the southern extreme of the distribution, Póvoa and Agudela (ME ecotype) show the lowest values not only for heterozygosity but also for the other parameters (Table 11). The highest percentage of polymorphic loci is observed in Muros (P_{95}) and Cangas (P_{99}), the highest *Ar* in Oia and Abelleira, while this last population presents the highest *Ar* and Tirán the highest *PAr*. Among the ME populations, Oia presents the highest diversity at all parameters except *PAr*; whereas among the ZS populations, Grove 2 tends to present higher diversity than Grove 1, except for *PAr* as well (Table 11).

When we compare the *L. fabalis* populations presenting the highest genetic diversity (Muros and Abelleira) with the *L. obtusata* populations, the first present similar or higher variability than the latter, independently of the parameters considered. Within *L. obtusata*, Cabo do Mundo shows higher genetic diversity than Moinhos for all parameters analyzed (Table 11).

Table 11. Genetic diversity in flat periwinkles' populations from the IP. *He*, expected heterozygosity; *Hnb*, non-biased heterozygosity; *Hobs*, observed heterozygosity; *A*, number of alleles per locus; *Ar*, allelic richness; *PAr*, private allelic richness; P_{95}, percentage of polymorphic loci using the 95% criterion; P_{99}, percentage of polymorphic loci using the 99% criterion. Values between brackets represent the standard deviation.

	Ecotype	**Species**	***He***	***Hnb***	***Hobs***	P_{95}	P_{99}	***A***	***Ar***	***PAr***
Abelleira	FI	*L. fabalis*	0.534 (0.315)	0.546 (0.322)	0.522 (0.331)	0.813	0.8750	5.8750	5.47	0.44
Muros	FI	*L. fabalis*	0.582 (0.250)	0.595 (0.256)	0.496 (0.254)	0.938	0.9375	5.5000	5.20	0.28
Grove 1	ZS	*L. fabalis*	0.470 (0.310)	0.481 (0.317)	0.464 (0.317)	0.813	0.8125	4.9375	4.65	0.07
Grove 2	ZS	*L. fabalis*	0.511 (0.312)	0.525 (0.284)	0.4686 (0.322)	0.867	0.8667	5.1333	5.13	0.06
Tirán	FI	*L. fabalis*	0.488 (0.325)	0.495 (0.330)	0.465 (0.346)	0.750	0.8125	5.8125	4.91	0.50
Cangas	FI	*L. fabalis*	0.501 (0.292)	0.513 (0.299)	0.458 (0.304)	0.875	1.0000	5.0000	4.80	0.13
Canido	FI	*L. fabalis*	0.479 (0.308)	0.489 (0.315)	0.458 (0.307)	0.813	0.8750	5.1875	4.82	0.17
Silleiro	ME	*L. fabalis*	0.494 (0.292)	0.504 (0.257)	0.475 (0.285)	0.867	0.8667	4.6000	4.51	0.17
Oia	ME	*L. fabalis*	0.499 (0.332)	0.511 (0.305)	0.501 (0.324)	0.867	0.9333	5.7333	5.47	0.15
Moinhos	Obt	*L. obtusata*	0.415 (0.348)	0.421 (0.311)	0.430 (0.335)	0.800	0.8667	3.8000	3.58	0.52
Póvoa	ME	*L. fabalis*	0.270 (0.246)	0.276 (0.252)	0.229 (0.229)	0.563	0.6875	2.1250	2.07	0.00
Agudela	ME	*L. fabalis*	0.354 (0.336)	0.359 (0.289)	0.349 (0.322)	0.667	0.7333	2.7333	2.99	0.06
Cabo do Mundo	Obt	*L. obtusata*	0.506 (0.262)	0.517 (0.223)	0.431 (0.251)	0.933	0.9333	4.2000	4.12	0.10

3.1.2. Genetic structure and phylogenetic relationships

The STRUCTURE results for *K*=2 (the best *K* according to Evanno's method, Figure 15A) show a clear separation between *L. fabalis* (grey) and *L. obtusata* (pink) individuals, with some putative hybrids between the two species (mainly in Cabo do Mundo) (Figure 16). Moinhos, where all males had a typical *L. obtusata's* penis, presents 100% membership to the pink cluster, while the majority of individuals from the *L. fabalis* populations present a high membership to the gray cluster.

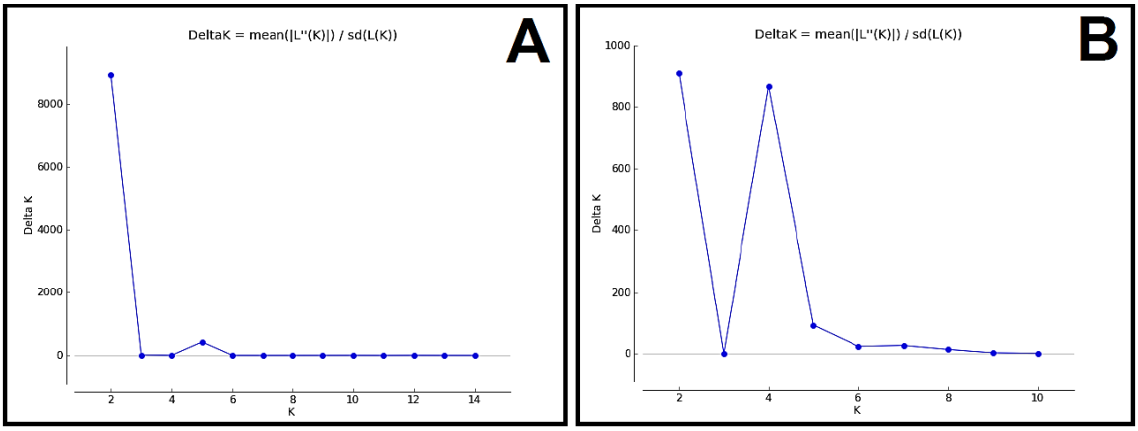

Figure 15. Delta-*K* values for the different number of clusters (*K*) following Evanno's method. A) For STRUCTURE runs with *L. fabalis* and *L. obtusata*; B) For STRUCTURE runs containing only *L. fabalis* individuals.

One individual from Abelleira (female), two from Grove 2 (one male and one female) and three from Cangas (one female and two males) were genetically assigned to *L. obtusata* (*Q*>0.80) (Figure 16). Although these females' species status could not be assessed using a diagnostic phenotypic trait, the two males present a penis morphology characteristic of *L. obtusata* but were included to assure the utility of these markers for the species distinction, even when no prior information (e.g. morphological, geographical) is included. Interestingly, many males from Cabo do Mundo, most of which possess a typical *L. obtusata* penis, exhibit an admixed genetic composition, suggesting genetic introgression (Figure 16).

After the exclusion of *L. obtusata* individuals and putatively misclassified/introgressed females, while the Evanno's method suggests *K*=2 (or *K*=4) as the number of clusters that better fit the data (Figure 15B), our empirical method (see Methods) rather points to *K*=8, which is compatible with the substructure of *L. fabalis* populations being mainly governed by geography rather than by ecology (i.e. according to ecotype classification) (Figure 17). For *K*=2, the Portuguese (Southern) populations (Póvoa and Agudela – ME) are separated from all Galician (Northern) ones. For *K*=3, apart from the Portuguese cluster, the Galician populations are split into a Southern (Ría de Vigo: Cangas and Tirán - FI, Silleiro and Oia – ME) and a Northern clade (Abelleira and Muros - FI, Grove 1 and Grove 2 - ZS). For *K*=4, the Southern Galician clade is further separated into Southernmost populations (Silleiro and Oia – ME) and Ría de Vigo populations (Cangas, Tirán and Canido - FI).

For *K*=5, the Northern Galician cluster is divided in two additional clades: Ría de Muros e Noia (Abelleira and Muros - FI) and Ría de Arousa (Grove 1 and Grove 2 - ZS) populations. For *K*=6, a split is observed between the Southernmost Galician populations, Oia and Silleiro (ME). The subsequent increase in the number of clusters (*K*) resulted in less prominent subdivisions; with a further separation within the FI populations of Ría de Vigo (Northern - Cangas and Tirán, and Southern - Canido) for *K*=7, as well as between Grove 1 and Grove 2 within Ría de Arousa for *K*=8.

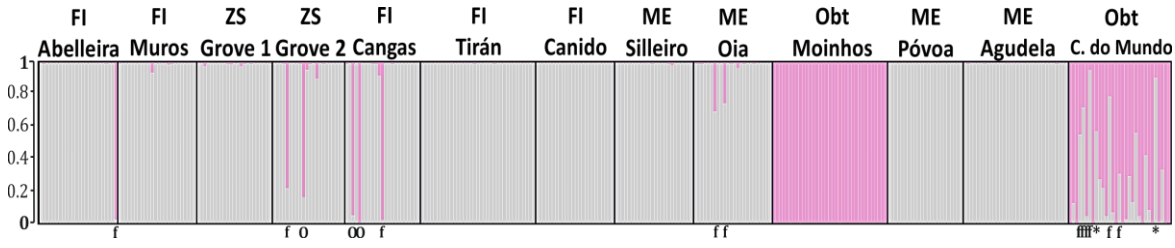

Figure 16. Membership of individuals to the clusters identified by the algorithm implemented in STRUCTURE for *K*=2. Membership is represented in the Y-axis scale: 1 corresponds to 100% membership to one of the two genetic clusters (gray bars - *L. fabalis*, pink bars – *L. obtusata*). Bars with both colors represent individuals with membership different from zero to both clusters and, most likely, admixed ancestry. On top, location name, together with the ecotype or species present in each site, is indicated. In the bottom, codes for individuals that do not match the most common cluster in the corresponding location: "O" indicates *L. obtusata* males; "*", *L. fabalis* males; and "f", females that present a genetic membership lower than 0.80 to the most common cluster.

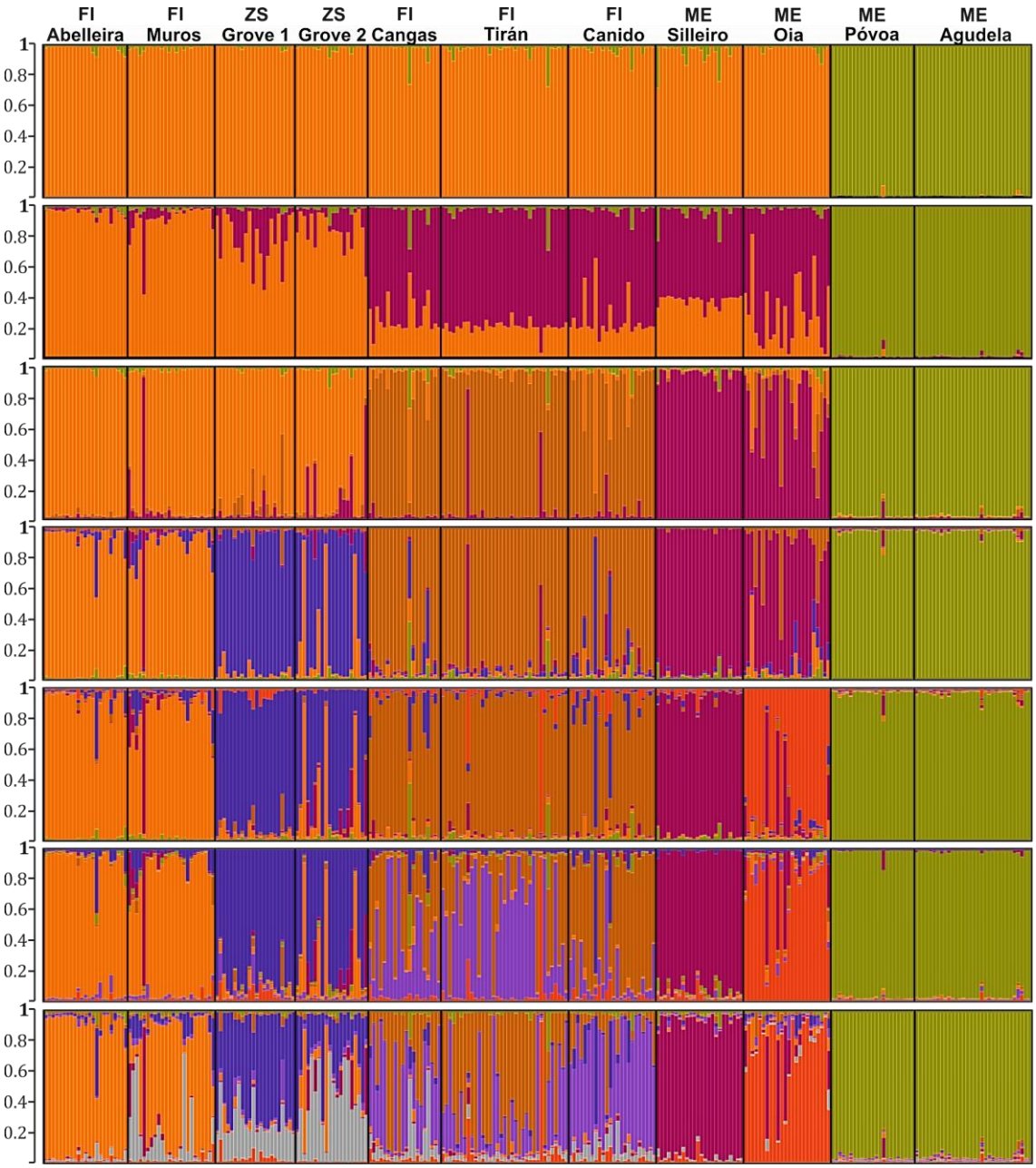

Figure 17. Membership of individuals to the clusters identified by the algorithm implemented in STRUCTURE from *K*=2 to *K*=8. *L. obtusata*'s individuals and putatively misclassified/introgressed females were removed from this analysis.

Mean F_{ST} is higher between species (0.4266) than between populations within species (Figure 18A), with the differentiation within *L. obtusata* being slightly higher than within *L. fabalis* (0.1633 vs. 0.1592, Figure 18A). Within *L. fabalis*, the differentiation between ecotypes is higher when the estimate involves the ME ecotype (0.2281 and 0.1922 vs. 0.0971) (Figure 18B). Mean F_{ST} within ME (i.e. between ME populations) is higher than within FI and ZS (0.2406 vs. 0.0633 and 0.0414, respectively), and even higher than between ecotypes, contrary to FI and

ZS, where intra-ecotype differentiation is lower than between them (0.0633 and 0.0414 vs. 0.0971) (Figure 18B).

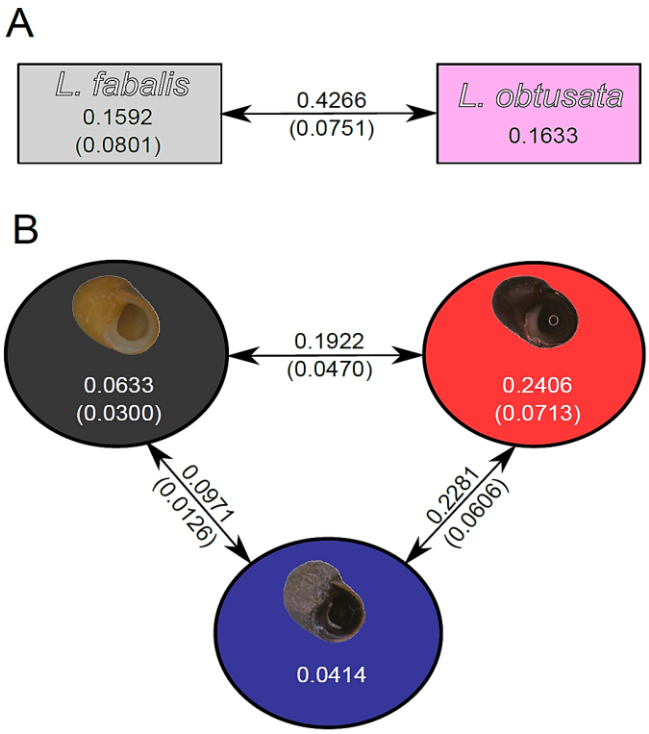

Figure 18. Representation of the F_{ST} values between and within species (A) and ecotypes (B). Values represent average F_{ST} over loci with the standard deviations in brackets. A) In grey all the *L. fabalis* populations, in pink *L. obtusata* from Moinhos and Cabo do Mundo. B) In black the FI ecotype, blue (ZS ecotype) and red (ME ecotype).

The Mantel test within *L. fabalis* revealed a non-significant association between genetic differentiation and geographic distance (i.e. IBD), except when the ME ecotype was removed from the dataset (Spearman rank correlation coefficient, ρ=0.44, p<0.05; Figure 19A). This pattern is robust to the differentiation measure used to perform the Mantel test, which is corroborated by the significant correlation observed between R_{ST} and F_{ST} (Spearman rank correlation coefficient, ρ=0.83, p<0.05; Figure 19B).

The population-based (Figure 20A) and individual-based (Figure 20B) NJ trees show similar results to the STRUCTURE analysis, both in terms of the clear split between *L. obtusata* and *L. fabalis*, the patterns of differentiation between *L. fabalis* populations, as well as the more intermediate position of Cabo do Mundo, suggesting introgressive hybridization between the two species. Furthermore, FI

individuals present a more widespread distribution in the tree than the other ecotypes, whereas some degree of connectivity among populations through migration and/or gene flow is suggested by the lack of 100% correspondence between the sampling location of individuals and their genetic position in the tree, also in agreement with the STRUCTURE results (Figure 17).

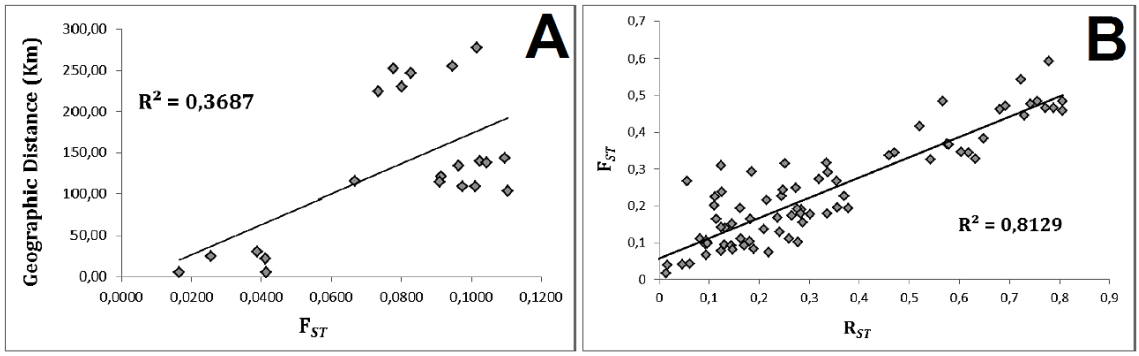

Figure 19. Genetic differentiation between *L. fabalis* populations. A) Plot of genetic differentiation (F_{ST}) against geographic distance after removing ME individuals (p=0.032). B) Plot of R_{ST} against F_{ST}.

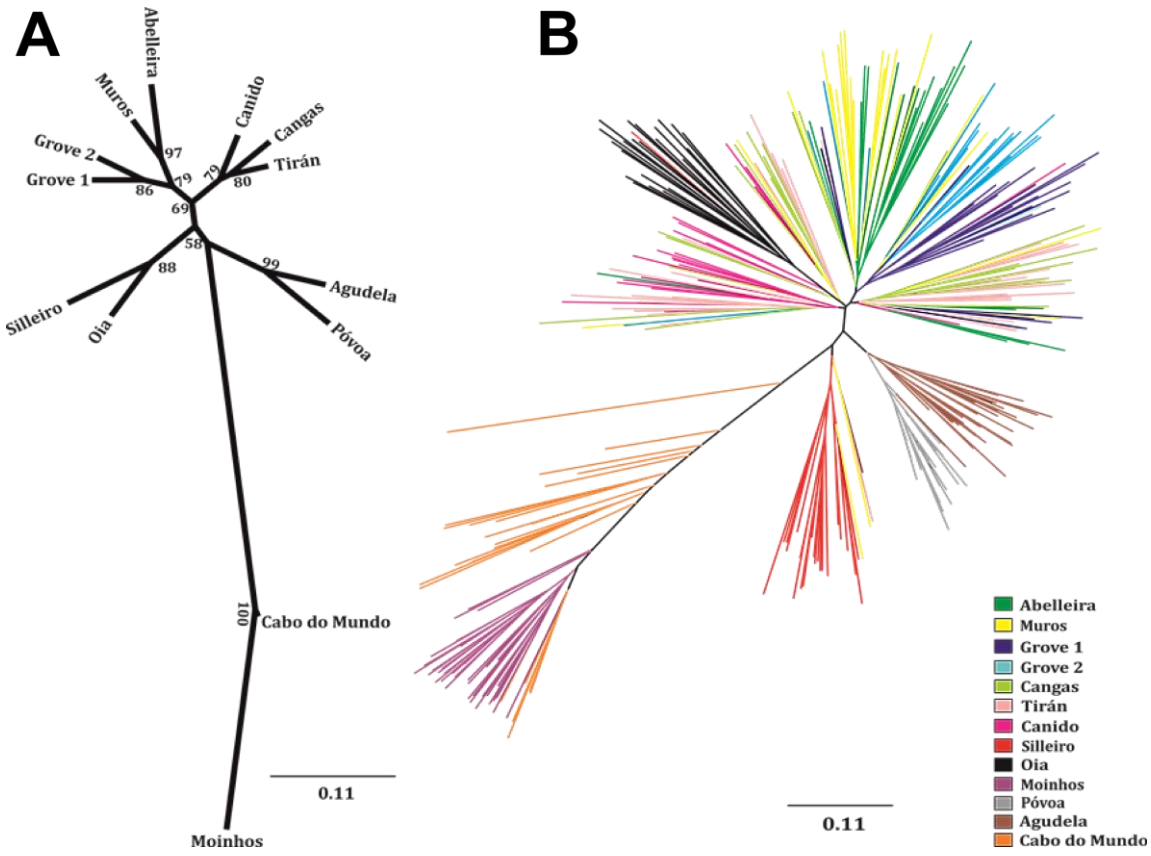

Figure 20. Unrooted NJ trees based on Nei's genetic distance. A) Population-based tree, with numbers at nodes indicating the bootstrap support (only those >50 are shown). B) Individual-based tree (each branch represents an individual, with a different color assigned to each population). Bootstrap supports for the individual-based tree are not shown, as they were generally low (<74).

3.2. Genetic analysis of L. fabalis ecotypes in Northern Europe based on AFLPs

3.2.1. Genetic diversity and genetic structure

A total of 681 polymorphic loci were analyzed in 379 individuals from 17 populations in Norway, Sweden and UK. The mean expected heterozygosity (*He*) across all populations was 0.2993, whereas the mean percentage of polymorphic loci (P_{95}) was 84.9%. Mean *He* per country was lower in Sweden (0.2837) and similar between Norway (0.3031) and UK (0.3053) (Table 12). An analogous trend was observed for the percentage of polymorphic loci, being lower in Sweden (83.9%) and similar between Norway (85.1%) and UK (85.3%).

Table 12. Genetic diversity for the studied populations and countries based on AFLP loci. *N*, number of individuals analyzed for each population; Number P_{95}, number of polymorphic loci using the 95% criterion; P_{95}, percentage of polymorphic loci using the 95% criterion; *He*, expected heterozygosity. Values between brackets represent the standard deviation. Estimates for each country are averaged between locations (in Syltonya only the most exposed site was included, Syl_ExpA). Mod-Exp stands for moderately-exposed.

Population	Country	Habitat	*N*	Number P_{95}	P_{95}	*He*
Hum_Shl	Norway	Sheltered	22	550	80.8	0.2913 (0.0066)
Hum_Exp	Norway	Mod-Exp.	23	572	84.0	0.2965 (0.0064)
Sel_Shl	Norway	Sheltered	25	594	87.2	0.3147 (0.0063)
Sel_Exp	Norway	Mod-Exp.	24	596	87.5	0.3064 (0.0061)
Syl_Shl	Norway	Sheltered	20	570	83.7	0.2984 (0.0063)
Syl_ExpA	Norway	Mod-Exp.	23	600	88.1	0.3237 (0.0061)
Syl_ExpB	Norway	Mod-Exp.	24	573	84.1	0.2909 (0.0064)
Norway	-	-	161	579	85.1	0.3031 (0.0124)
Lok_Shl	Sweden	Sheltered	24	573	84.1	0.2790 (0.0062)
Lok_Exp	Sweden	Mod-Exp.	23	571	83.8	0.2800 (0.0061)
Urs_Shl	Sweden	Sheltered	23	565	83.0	0.2792 (0.0063)
Urs_Exp*	Sweden	Mod-Exp. *	21	576	84.6	0.2965 (0.0063)
Sweden	-	-	91	571	83.9	0.2837 (0.0086)
AngN_Shl	UK	Sheltered	21	569	83.6	0.2995 (0.0060)
AngN_Exp	UK	Mod-Exp.	19	569	83.6	0.3012 (0.0060)
AngN_Int	UK	Intermediate	24	593	87.1	0.3017 (0.0061)
AngN_Unk	UK	Unknown	24	602	88.4	0.3101 (0.0059)
AngS_Shl	UK	Sheltered	17	562	82.5	0.3091 (0.0063)
AngS_Exp	UK	Mod-Exp.	22	590	86.6	0.3101 (0.0062)
UK	-	-	127	581	85.3	0.3053 (0.0049)

*Despite previous information that this site was moderately-exposed, the observed high density of *Ascophyllum* spp. suggests that it is rather sheltered.

When the populations were grouped according to wave-exposure, the analysis revealed higher P_{95} and *He* in populations living in exposed shores compared to sheltered ones (88.1% vs. 86.2% and 0.30422 vs. 0.29787, respectively), a pattern that maintains (mainly *He*) even when we compare exposed and sheltered sites within each location, separately (Table 12).

Pairwise F_{ST} between countries revealed UK as the most differentiated, particularly in respect to Sweden (highest $F_{ST,}$ 0.0612 vs. 0.0330 and 0.0392) (Figure 21A). In Norway (Figure 21B), the level of ecotype differentiation was more heterogeneous between locations, with Hummelsund presenting the highest differentiation (~0.07) and Sele the lowest (~0.05). In UK (Figure 21C), ecotype differentiation was similar between Southern and Northern Anglesey (~0.04), with values similar to Sele (Norway). Contrarily, the lowest F_{ST} between ecotypes was observed in Sweden (Figure 21D) (see Discussion).

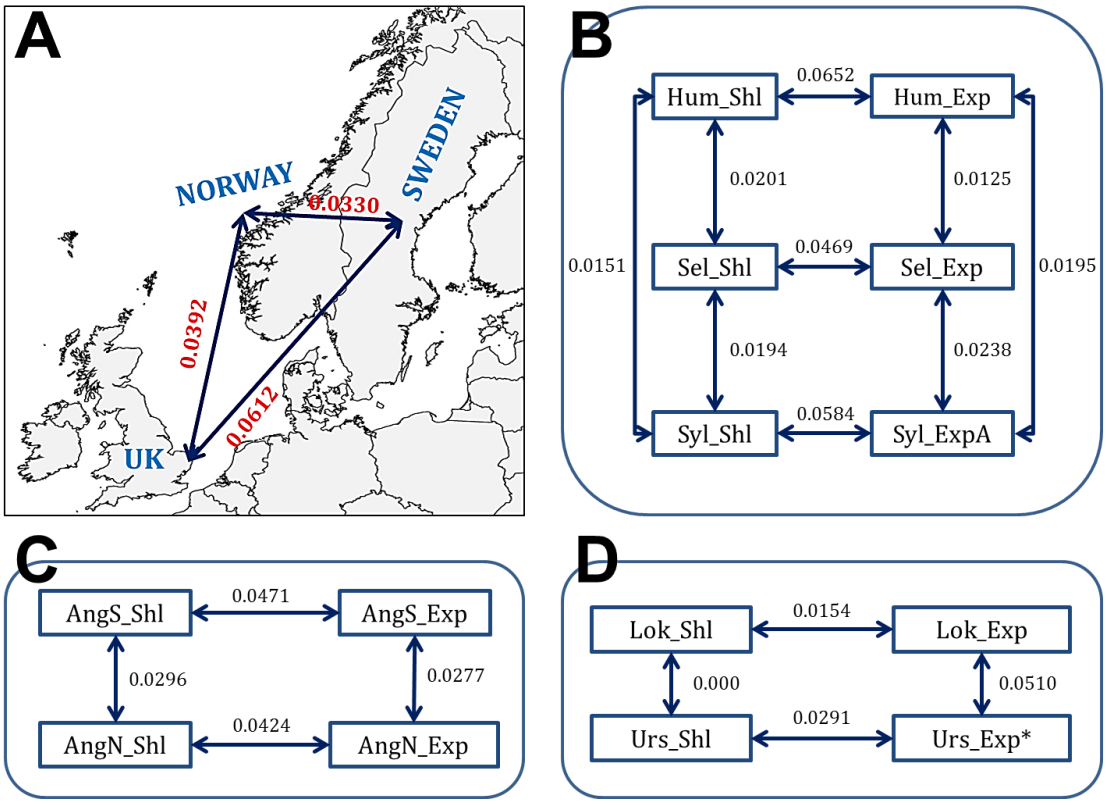

Figure 21. Representation of F_{ST} values between countries (A) and between populations within country (B, C, D) based on 681 AFLP loci. Information about the populations from B) Norway, C) UK and D) Sweden are available in Table 12. *Despite previous information that this site was moderately-exposed, the observed high density of *Ascophyllum* spp. suggests that it is rather sheltered. Sites with intermediate or unknown exposure were excluded from the analysis. In Syltonya, only the most exposed site (Syl_ExpA) was included.

3.2.2. Detection of outlier loci

The total number of outliers between all ecotype comparisons detected by MCHEZA was 138 (vs. 543 nonoutliers), but it considerably decreased (43) after the FDR correction, with no outliers in Sweden (Table 13). Although BAYESCAN detected less outliers (19), the observed trend across locations was similar to the

results of MCHEZA. The analysis with MCHEZA detected from 11 to 60 outlier loci in the comparisons between ecotypes within each locality, with Sweden presenting less outliers, in agreement with the lower genetic differentiation shown before.

Table 13. Summary of the outlier analysis. For each location, the number of outlier loci detected with MCHEZA and BAYESCAN when comparing the LM and SS ecotypes is presented. The MCHEZA outliers that remained after the false discovery rate (FDR) correction (alpha=0.1) are also indicated. The total number of different outliers for each country is also shown.

	Comparison	MCHEZA Outliers	MCHEZA Outliers (FDR)	BAYESCAN Outliers
	Hummelsund	60	24	8
	Sele	36	8	3
	Syltonya	39	7	4
Total Norway		**74**	**28**	**14**
	Lokholmen	24	0	0
	Ursholmen	11	0	0
Total Sweden		**35**	**0**	**0**
	Anglesey North	46	6	2
	Anglesey South	42	11	3
Total UK		**74**	**15**	**5**

The percentage of shared outliers between the three countries is 25% (34% between UK and Sweden; 37% between Norway and Sweden; and 43% between Norway and UK). Within each country, the percentage of shared outliers between locations is higher in Norway (47%) than in the UK (33%).

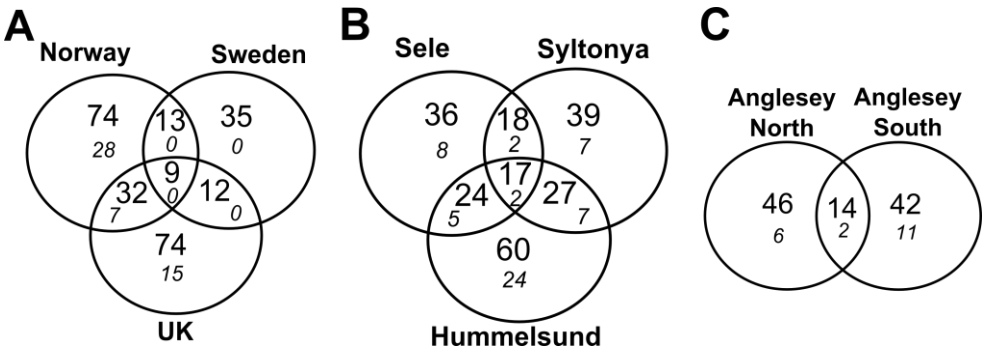

Figure 22. Representation of the number of outlier loci detected with MCHEZA and the number of outliers overlapping at different geographical scales. A) Between countries; B) between Norwegian locations and C) between British locations. The number of MCHEZA outlier loci with 0.1 FDR is displayed in italics.

3.2.3. Genetic structure of outlier and nonoutlier loci

Clusters detected by STRUCTURE varied with the dataset used. The nonoutlier loci revealed *K*=2 as the best fit, with an apparent division between Norway-Sweden (red color) and UK (green color) (top panel, Figure 23). The outlier loci also rendered *K*=2, with a separation between exposed (red color) and sheltered (green color) sites (bottom panel, Figure 23). Apparent exceptions occurred in Ursholmen, where the genetic constitution of the exposed population presents a sheltered genetic background, and in Anglesey North, where the intermediate and unknown populations show an admixed composition.

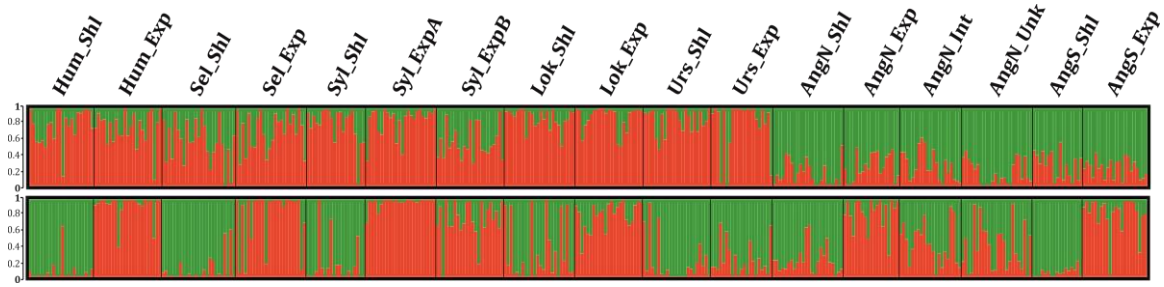

Figure 23. Membership of individuals to the clusters identified by the algorithm implemented in STRUCTURE (*K*=2). Using the 'nonoutlier' dataset (top panel) and the 'outlier' dataset (bottom panel). Location codes are presented above and are valid for both plots. Each color represents a genetic cluster. Membership is represented in the Y-axis scale: 1 corresponds to 100% membership to one of the two genetic clusters.

Independent STRUCTURE runs using the outlier dataset for each country revealed the same general pattern, with a clear separation between sheltered populations (green color) and exposed populations (red color) except for the two cases mentioned above (Figure 24). Independent runs for each region were also performed using the nonoutlier dataset for comparison, but no clear substructure was found within each country that can be associated with geography or ecotypes for *K*=2, with individuals of the two clusters present in all locations.

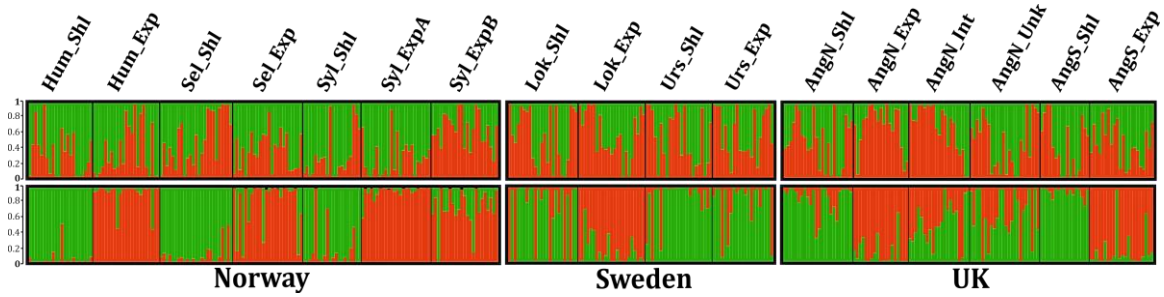

Figure 24. Membership of individuals to the clusters identified by the algorithm implemented in STRUCTURE for each country separately. *K*=2 is shown for comparison between the nonoutlier (top panel) and the outlier (bottom panel) datasets. Location codes are presented above whereas countries below the plots.

Discussion

The group of flat periwinkles (*L. fabalis* and *L. obtusata*) constitutes a system with great potential for investigating divergence in progress along the speciation continuum that has been largely overlooked. Here I present the first study on these species comprising populations from very distant regions of their distribution (ecotypes from Iberian Peninsula and Northern Europe) and combining morphological and genetic analyses.

1. Diversification of flat periwinkles in the Iberian Peninsula

The detailed information on the distribution of flat periwinkles in the North of Portugal, as well as in the South of Galicia, collected during this project was not only crucial to select suitable sampling sites for the subsequent genetic and morphological characterization of the Iberian populations performed here, but also to provide important data that will certainly facilitate sampling efforts in future evolutionary studies on these species. For example, to my knowledge this is the first work describing the presence of the *L. fabalis* ME ecotype in Portugal, where it tends to be more common than the FI ecotype, contrary to the pattern observed in Galician Rías (Rolán & Templado, 1987; Rolán-Alvarez et al., 1995). In the future, it would be interesting to investigate if this pattern results from distinct ecological conditions in the two regions (e.g. wave-exposure), different evolutionary history of the populations (e.g. different refugia), or a combination of multiple factors.

Previous genetic studies in flat periwinkles were based on a limited number of markers originally developed for other littorinids (Schmidt et al., 2007; Kemppainen et al., 2009; McInerney et al., 2009), thus presenting problems related with their cross-amplification (e.g. null alleles; Kemppainen et al., 2009). The battery of flat periwinkle specific microsatellite loci developed here represents a new and powerful molecular tool to assess genetic variation and differentiation between populations of these species, which can have numerous additional applications (e.g. effects of multiple paternity on the demographic history; Rafajlovic et al., 2013) besides the topic of this study.

The analysis of these microsatellite loci in the IP showed a clear separation between *L. fabalis* and *L. obtusata* (Figure 16), in almost perfect association with differences in penis morphology, showcasing the value of these loci for species discrimination, at least in the IP. Furthermore, it allows overcoming previous limitations derived from the fact that only adult males could be unambiguously assigned to species based on morphology. Since the species status of females (as well as of juveniles from either sex) can now be assessed with this new battery of markers, they can be incorporated in these kind of analyses, avoiding sex-bias problems in population genetics inference and diminishing sample size constrains.

Shell morphometric analysis revealed size and shape differences between *L. obtusata* and *L. fabalis*, although some overlap was observed in terms of shape (Figure 13). Although these phenotypic differences could result from adaptive divergence (e.g. crab predation), the contribution of other factors (e.g. genetic drift) cannot be excluded, highlighting the need for further studies to evaluate these hypotheses. Nonetheless, it is noteworthy that the GM protocol developed here indeed represents a valuable approach for the phenotypic characterization of flat periwinkles, with different potential applications in evolutionary biology (e.g. QTL analysis, phenotypic plasticity).

The admixture of species-specific traits found in Cabo do Mundo was confirmed at the genetic level, with the detection of extensive hybridization between the two species (Figure 16). Although this process was suggested before as a possible cause of the mtDNA shared variation between them in populations from NE, preference was given to the hypothesis of prevalent incomplete lineage sorting (Kemppainen et al., 2009). Moreover, the small number of nuclear markers previously available (4 microsatellite loci), together with the significant amount of null alleles detected, prevented an accurate assessment of the level of hybridization (Kemppainen et al., 2009). Here, the first strong evidence for introgressive hybridization between *L. fabalis* and *L. obtusata* at the nuclear genome (in the IP) was presented. The genotyping of this battery of microsatellites should be extended to populations from NE in order to evaluate the occurrence of introgressive hybridization between

flat periwinkles in this region. Episodes of hybridization between other sister species within the genus have also been detected (e.g. between *L. saxatilis* and *L. arcana*; Mikhailova et al., 2009), but how hybridization influences the evolution of these species, for instance through adaptive introgression as described for other systems (e.g. butterflies, Salazar et al., 2010, reviewed by Hedrick, 2013), is still unknown.

The substantial differences in terms of penis morphology between *L. fabalis* and *L. obtusata* have been the basis to propose that pre-zygotic barriers could be involved in their divergence (Reid, 1996), including a possible role for reinforcement (Hollander et al., 2013). However, the observed admixture based on morphological and genetic data shows that reproductive isolation is not complete. Therefore, other type of barriers (e.g. postzygotic) could explain the different levels of hybridization and introgression between the two species across locations. The lack of complete reproductive isolation suggests that these sister species could represent "late stages" of the speciation continuum (Hendry, 2009). Given that the two species occupy different habitats, ecologically-based natural selection could be important as a barrier to gene flow in flat periwinkles (Schluter, 2000). Cabo do Mundo could be used in future studies as a natural laboratory to quantify the rate of hybridization, to investigate the reproductive barriers between these sister species and to uncover whether this hybridization is adaptive or not. In general, "late stages" of ecological speciation are not as well characterized as earlier stages (Nosil, 2012) and, consequently, systems like this one are attracting a growing interest among the research community.

Within *L. fabalis,* the genetic structure of populations revealed a pattern that seems to be more influenced by geography (distance) than by ecology (i.e. ecotype classification), with a clear first split between Northern Portugal and Galicia (Figure 17 and 20), and then among the different Galician Rías and between these and the exposed shores of Silleiro and Oia (ME ecotype). Indeed, Río Minho, which divides Northern Portugal and Galicia, could represent a barrier to gene flow as observed between populations of *Fucus vesiculosus* from both sides of the river (Zardi et al., 2013), which could also be the case for *L. fabalis*. Although isolation by

distance (IBD, Mantel test) was not significant, this could be (at least partially) explained by the higher differentiation between populations when the ME ecotype is involved, both in intra- and inter-ecotype comparisons (Table 10, Figure 18). The extreme wave exposure faced by the ME ecotype could cause higher mortality, leading to stronger genetic drift, possible bottlenecks and/or extinctions followed by re-colonization, resulting in higher genetic differentiation. Indeed, this is supported by the observation of lower density of snails in some ME populations (Silleiro and Oia), particularly in certain periods of the year. On the other hand, F_{ST} estimates within the ZS ecotype were the lowest, pointing to weaker genetic drift or higher gene flow between ZS populations than in the other ecotypes. However, this should be interpreted with caution since only two geographically close ZS populations (the only ones described so far) were used. Moreover, due to constraints associated with the distribution of the ecotypes (e.g. the restricted geographic distribution of the ZS ecotype), it is difficult to separate the effects of geography and ecology on the observed differentiation patterns.

In addition to the contribution of geographic and stochastic factors, it is likely that natural selection also plays a role in ecotype differentiation. In particular, the smaller size of the ME ecotype (Table 8) combined with its peculiar habitat (a very different host algae living in heavy wave-exposure shores) could have promoted its distinctiveness and the evolution of additional barriers to gene flow in respect to the other ecotypes; like in the case of *L. saxatilis*, where size is associated with different habitats and the presence of size-related assortative mating represents an important barrier to gene flow (Rolán-Alvarez, 2007; Butlin et al., 2014). However, this hypothesis has to be tested in *L. fabalis*, for example by performing mating trials in the laboratory and reciprocal transplants in the field to test the effect of natural selection in phenotypic divergence and reproductive isolation between ecotypes (reviewed in Nosil, 2012). It would be important to study these ecotypes by means of genome scans (e.g. AFLP loci), to identify the outliers associated with adaptive divergence, as in *L. saxatilis* (Galindo et al., 2009; 2013) and many other organisms (reviewed in Nosil et al., 2009b).

Interestingly, contrary to the pattern observed in NE (and in *L. saxatilis*), the Iberian ecotypes do not show detectable differences in shape, suggesting that shell shape plays a minor role in their differentiation. However, this system is unique in the sense that the color of the different ecotypes mimics their distinct host algae (see Introduction, Figure 5). Shell color (not addressed in this work) has been shown to be under selection due to predation in other populations of *L. fabalis* (Reimchen, 1979). Therefore, future studies are needed to unravel the evolutionary forces acting on this trait, as well as its interaction with size and shape, in an effort to understand if these phenotypic differences result from the action of natural selection, alone or in combination with other mechanisms (e.g. genetic drift, phenotypic plasticity), and to determine the contribution of these processes to the diversification of *L. fabalis* in the IP.

2. Diversification of flat periwinkles in Northern Europe

Although *L. fabalis* ecotypes in NE have been the target of various studies (Tatarenkov & Johannesson, 1998, 1999; Kemppainen et al., 2005, 2009, 2011), this work presents a thorough morphological characterization and the first genome scan performed in these ecotypes, with a similar experimental design to that employed in *L. saxatilis* by Butlin et al. (2014) to test for parallel speciation.

The GM approach implemented here confirmed the previously suggested morphological differences (in terms of shell size and shape) between the LM and SS ecotypes (Johannesson & Mikhailova, 2004; Kemppainen et al., 2009), which are relatively constant across the three studied countries (Table 9, Figure 14). The repeated phenotypic divergence detected here suggests a role of divergent natural selection as proposed for *L. saxatilis* (Johannesson, 2003; Rolán-Alvarez, 2007) and other model systems of ecological speciation (Rundle et al., 2000; Nosil et al., 2002), though the size trend in these ecotypes from NE is the opposite to what is commonly found in other intertidal gastropod species (Kemppainen et al., 2005) and even in the *L. fabalis* ecotypes from the IP (see above). In NE populations, the increased risk of dislodgment in more exposed habitats has been proposed to

create a selective pressure for a larger size because these individuals are able to more effectively withstand crab predation when they fall off their host algae (Kemppainen et al., 2005). However, the effect of phenotypic plasticity cannot be ruled out, highlighting the need for additional experiments such as the ones performed in *L. saxatilis* (Hollander et al., 2006; Hollander & Butlin, 2010).

Although natural selection has already been claimed to be involved in the genetic divergence between *L. fabalis* ecotypes in NE (Tatarenkov & Johannesson, 1998; Kemppainen et al., 2011), this is the first time that signatures of natural selection are investigated at a genome-wide scale. The AFLP genome scan applied here revealed 11 to 60 outlier loci (MCHEZA) by locality (~5% of the genome) (Table 13, Figure 22), a result very similar to analogous studies in *L. saxatilis* (Wilding et al., 2001; Galindo et al., 2009, 2013; Butlin et al., 2014) as well as in a wide range of other organisms (reviewed in Nosil et al., 2009b). The lowest number of outliers (11) was detected in Ursholmen (Sweden), which was not surprising given that one of the sampling sites had been initially misclassified as truly exposed. When comparing localities within each country (only in Norway and UK, since in Sweden the localities did not represent the same environmental cline), 47% of outliers were shared among the three Norwegian locations and 33% between the two British locations, in spite of the larger geographical distance between the former (~100 Km) respect to the later (~60 Km).

The proportion of outliers shared between countries was relatively high (34-43% of the total) (Figure 22), but this has to be interpreted with caution because of the unequal number of locations inspected in each country (three in Norway, two in UK and one truly exposed-sheltered locality in Sweden). This proportion is greater than the one observed for *L. saxatilis*' ecotypes in Galicia for AFLPs (9-21%) (Galindo et al., 2013), despite the different geographic scales of the two studies (i.e. here, the distance between locations from different countries is 1000s Km, whereas in Galindo et al. (2013) locations are less than 100 Km apart). Furthermore, a recent transcriptome scan (RNA-seq) performed in *L. saxatilis* showed a lower level of shared outliers (~15%) among similar regions (Sweden and UK), though the methodology is not entirely equivalent to the one implemented here (Westram et

al., 2014). In general, it seems that a relevant proportion of outlier loci (a proxy for adaptive variation) are shared over a larger geographic range in NE flat periwinkles.

Taking into account that the genomic position of the different AFLP loci is currently unknown, the higher percentage of sharing observed in this work could be explained by linkage (i.e. not complete independence) among several AFLP outlier loci, for instance due to genetic hitchhiking or their location within an inversion (Nosil et al., 2009a; Faria & Navarro, 2010). On the other hand, such level of sharing could also be explained by a more recent history of *L. fabalis* ecotypes in NE or gene flow over large scales (e.g. evolution in concert) (Johannesson et al., 2010); in agreement with the lower differentiation between the flat periwinkle populations from different countries analyzed here ($0.03<F_{ST}<0.06$), when compared with *L. saxatilis* (F_{ST}=0.11, Sweden vs. UK; Westram et al., 2014).

Remarkably, a previous study suggested that one arginine kinase haplotype (or a SNP linked to this gene), under positive selection in sheltered populations of *L. fabalis* from NE, is shared across locations over a geographical scale similar to the one used here (Kemppainen et al., 2010), but the origin of that adaptive variation (e.g. gene flow, ancestral polymorphism, de novo mutations) is still under debate (Johannesson et al., 2010; Faria et al., 2014). Thus, a follow up sequencing study of the outlier loci detected here (nine shared across all countries) is needed to distinguish between these different evolutionary scenarios (see Wood et al., 2008). Nevertheless, it seems reasonable to think that *L. fabalis* ecotypes in NE are diverging due to ecologically-based natural selection associated with their different habitats and that at least part of these outlier loci represent regions of the genome truly affected by this process (while others might also exist). Furthermore, the LM and SS ecotypes could represent an early stage in ecological speciation, since gene flow between them is most likely a factor at play given the relative small distances between their habitats. This idea is also supported by the lack of genetic distinctiveness between LM and SS ecotypes when neutral genetic variation is analyzed (see below, Figure 24).

In addition, the data generated by this genome scan was used to investigate the genetic structure of *L. fabalis* populations in NE. The AFLP loci, based on all comparisons between contrasting ecotypes at each sampled location, were partitioned into an outlier (138 loci) and a nonoutlier (543 loci) dataset. The outlier loci revealed a clear split between LM and SS populations, which appear as two well-defined clusters independently of their region of origin (Figure 23). This result is similar to that found in *L. saxatilis* (Wilding et al., 2001; Galindo et al., 2013) and in other species (e.g. beetles, Egan et al., 2008; walking stick insects, Nosil et al., 2008) where divergence between contrasting ecotypes has been claimed to be generated/maintained by natural selection. However, a link between the outlier loci detected here and the distinct selective pressures faced by each ecotype needs to be uncovered before concluding that ecological speciation is ongoing in this system. Meanwhile, the nonoutlier loci rendered a geographic clustering of the sampled populations, which are divided into two clusters (Figure 23), one composed by individuals from Norway and Sweden, and another one by individuals from UK. The higher differentiation of UK relative to Sweden and Norway is also supported by overall F_{ST} differentiation between countries (Figure 21) and correlates well with geographic distance among them, although it can also be explained by the re-colonization of Scandinavia after the last glacial maximum from a refuge close to the English Channel, as suggested for *L. saxatilis* (Panova et al., 2011).

The contrasting clustering of these *L. fabalis* populations when based on neutral markers (i.e. nonoutliers - by geography) vs. putative adaptive markers (i.e. outliers - by ecology) suggest a parallel origin of the ecotypes, i.e. their independent evolution at least twice, in UK and Scandinavia. Nonetheless, gene flow among ecotypes within each region could also render a similar pattern if each ecotype had a single origin followed by secondary contact between them in both UK and Scandinavia (Faria et al., 2014). Thus, it would be interesting to apply a model-based approach (e.g. Approximate Bayesian Computation - ABC), such as the one performed in *L. saxatilis* by Butlin et al. (2014), in order to distinguish between parallel vs. single origin of the *L. fabalis* ecotypes in NE.

In any case, the origin of the genes responsible for the repeated phenotypic differentiation (same phenotype-habitat association) found between LM and SS ecotypes across the countries analyzed here is currently unknown, and the role of that differentiation in the evolution of reproductive isolation is also uncertain. Therefore, further studies should focus on these questions to determine if *L. fabalis* actually represents an example of parallel speciation (Nosil, 2012).

Conclusions

The goal of this study was to improve our knowledge about the distribution of phenotypic and genetic variation of *L. fabalis* across populations from the IP and NE, as well as about the process of divergence between *L. fabalis* and *L. obtusata*.

For the first time, unequivocal evidence for introgressive hybridization between these two species was presented, demonstrating the power of the new tools developed here (GM protocol, microsatellite loci) for species discrimination and population characterization.

The genetic structure of Iberian *L. fabalis* populations (based on microsatellite loci) revealed a preponderant role of geography in differentiation, together with other stochastic processes (e.g. genetic drift); whereas the phenotypic divergence between ecotypes points to a possible role of natural selection in their diversification, which needs to be explored in future studies by means of genome scans.

In NE, the implemented genome scan revealed a relatively high proportion of shared outliers (putative adaptive variation) between *L. fabalis* ecotypes across countries, which combined with their repeated phenotypic divergence (as corroborated here) supports that the LM and SS ecotypes are likely diverging under the influence of natural selection in the face of moderate gene flow.

The outcomes of this project represent the first steps towards establishing the flat periwinkles as a model system to investigate how reproductive barriers evolve and interact with each other across the speciation continuum (from ecotypes to species: LM vs. SS – *L. fabalis* vs. *L. obtusata*), contributing to move the field forward. Finally, the developments achieved in this study open up the possibility of performing comparisons with the *L. saxatilis* system in order to gain a better understanding of the mechanisms operating at different phylogenetic depths of diversification in marine intertidal gastropods.

References

Antao, T.; Beaumont, M. (2011) Mcheza: A workbench to detect selection using dominant markers. Bioinformatics 27(12), 1717-1718.

Beaumont M.A.; Nichols R.A. (1996) Evaluating loci for use in the genetic analyses of population structure. Proceedings of the Royal Society of London Series B 263, 1619–1626.

Benjamini, Y.; Hochberg, Y. (1995) Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 57(1), 289-300.

Berner, D.; Roesti, M.; Hendry, A.P.; Salzburger, W. (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. Molecular Ecology 19, 4963–4978.

Belkhir, K.; Borsa, P.; Goudet, J.; Chikhi, L.; Raufaste, N.; Bonhomme, F. (1996) GENETIX 4.03, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Univiversité de Montpellier II, Montpellier, France.

Boughman, J.; Rundle, H.; Schluter, D. (2005) Parallel evolution of sexual isolation in sticklebacks. Evolution 59(2), 361–373.

Butlin, R. (2010) Population genomics and speciation. Genetica 138(4), 409-418.

Butlin, R.; Saura, M.; Charrier, G.; Jackson, B.; André, C.; Caballero, A.; Coyne, J.; Galindo, J.; Grahame, J.; Hollander, J.; Kemppainen, P.; Martínez-Fernández, M.; Panova, M.; Quesada, H.; Johannesson, K.; Rolán-Alvarez, E. (2014) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution 68(4), 935–949.

Carvajal-Rodríguez, A.; Conde-Padín, P.; Rolán-Alavarez, E. (2005) Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of *Littorina saxatilis*. Journal of Molluscan Studies 71, 313–318.

Coyne, J.A.; Orr, H.A. (2004) Speciation. Sunderland, MA: Sinauer Associates.

Earl, A.; vonHoldt, B. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources 4(2), 359-361.

Egan, S.P.; Nosil, P.; Funk, D. J. (2008). Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution, 62(5), 1162-1181.

Ehrich, D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes 6, 603-604.

Elias, M.; Faria, R.; Gompert, Z.; Hendry, A. (2012) Factors influencing progress toward ecological speciation: a special issue. International Journal of Ecology.

Evanno, G.M.; Regnaut, S.; Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8), 2611-2620.

Falush, D.; Stephens, M.; Pritchard, J. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7(4), 574-578.

Faria, R.; Navarro, A. (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends in Ecology & Evolution 25(11), 660-669.

Faria, R.; Renaut, S.; Galindo, J.; Pinho, C.; Melo-Ferreira, J.; Melo, M.; Jones, F.; Salzburger, W.; Schluter, D.; Butlin, R. (2014) Advances in Ecological Speciation: an integrative approach. Molecular Ecology 23, 513-521.

Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35, 124–138.

Foll, M.; Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180(2), 977-993.

Funk, D. J. (2012) Of "Host Forms" and Host Races: Terminological issues in Ecological speciation. International Journal of Ecology.

Galindo, J.; Morán P.; Rolán-Alvarez, E. (2009) Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. Molecular Ecology 18, 919–930.

Galindo, J.; Martínez-Fernández, M.; Rodríguez-Ramilo, S.; Rolán-Alvarez, E. (2013). The role of local ecology during hybridization at the initial stages of ecological speciation in a marine snail. Journal of Evolutionary Biology 26(7), 1472-1487.

Goudet, J. (1995). Fstat version 1.2: a computer program to calculate Fstatistics. Journal of Heredity 86(6), 485-486.

Hedrick, P.W. (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular ecology 22(18), 4606-4618.

Hendry, A.P.; Bolnick, D.I.; Berner, D.; Peichel C.L. (2009) Along the speciation continuum in sticklebacks. Journal of Fish Biology 75, 2000-2036.

Hollander, J.; Collyer, M.; Adams, D.; Johannesson, K. (2006). Phenotypic plasticity in two marine snails: constraints superseding life history. Journal of Evolutionary Biology 19(6), 1861-1872.

Hollander, J.; Butlin, R. (2010). The adaptive value of phenotypic plasticity in two ecotypes of a marine gastropod. BMC Evolutionary Biology 10(1), 333.

Hollander, J.; Smadja, M.; Butlin, R.; Reid, D. (2013) Genital divergence in sympatric sister snails. Journal of Evolutionary Biology 26(1), 210-215.

Jakobsson, M.; Rosenberg, N. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14), 1801-1806.

Johannesson, K. (2003) Evolution in *Littorina*: Ecology matters. Journal of Sea Research 49, 107-117.

Johannesson, K.; Mikhailova, N. (2004) Habitat-related genetic substructuring in a marine snail (*Littorina fabalis*) involving a tight link between an allozyme and a DNA locus. Biological Journal of the Linnean Society 81, 301–306.

Johannesson, K.; Panova, M.; Kemppainen, P.; André, C.; Rolán-Alvarez, E.; Butlin, R. (2010). Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 365(1547), 1735-1747.

Kalinowski, S. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5(1), 187-189.

Kaliontzopoulou, A. (2011) Geometric morphometrics in herpetology: modern tools for enhancing the study of morphological variation in amphibians and reptiles. Basic and Applied Herpetology 25, 5-32.

Kemppainen, P.; Nes, S.; Ceder, C.; Johannesson. K. (2005) Refuge function of marine algae complicates selection in an intertidal snail. Oecologia 143, 402–411.

Kemppainen, P.; Panova, M.; Hollander, J.; Johannesson, K. (2009) Complete lack of mitochondrial divergence between two species of NE Atlantic marine intertidal gastropods. Journal of Evolutionary Biology 22, 2000-2011.

Kemppainen, P.; Lindskog, T.; Butlin, R.; Johannesson, K. (2011) Intron sequences of arginine kinase in an intertidal snail suggest an ecotype-specific selective sweep and a gene duplication. Heredity 106, 808–816.

Lynch, M; Milligan, B.G. (1994) Analysis of population genetic structure with RAPD markers. Molecular Ecology 3, 91-99.

Malausa, T.; Gilles, A. et al. (2011) High-throughput microsatellite isolation through 454 GS-FLX titanium pyrosequencing of enriched DNA libraries. Molecular Ecology Resources 11(4), 638-644.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27(2 Part 1), 209-220.

McInerney, C.E.; Allcock, A.L.; Johnson, M.P.; Prodöhl, P.A. (2009) Characterization of polymorphic microsatellites for the rough periwinkle gastropod, Littorina saxatilis (Olivi, 1792) and their cross-amplification in four congeners. Conservation genetics, 10(6), 1989-1992.

McKinnon, J.; Mori S.; Blackman, B.; David, L.; Kingsley, D.; Jamieson, L.; Chou, J.; Schluter, D. (2004) Evidence for ecology's role in speciation. Nature 429, 294-298.

Mikhailova, N.; Gracheva, Y.; Backeljau, T.; Granovitch, A. (2009) A potential species-specific molecular marker suggests interspecific hybridization between sibling species *Littorina arcana* and *L. saxatilis* (Mollusca, Caenogastropoda) in natural populations. Genetica 137(3), 333-340.

Nei, M.; Tajima, F.; Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. Journal of Molecular Evolution 19(2), 153-170.

Nosil, P. (2007) Divergent host-plant adaptation and reproductive isolation between ecotypes of *Timema cristinae*. American Naturalist 169, 151-162.

Nosil, P.; Egan, S.P.; Funk, D.J. (2008) Heterogeneous genomic differentiation between walking‐stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution, 62(2), 316-336.

Nosil, P.; Harmon, L.; Seehausen, O. (2009a) Ecological explanations for (incomplete) speciation. Trends in Ecology and Evolution 24(3), 145-156.

Nosil, P.; Funk, D.J.; Ortiz-Barrientos, D. (2009b) Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18(3), 375-402.

Nosil, P. (2012) Ecological speciation. Oxford, Oxford University Press.

Oosterhout, V.; Hutchinson, W.; Wills, D.; Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4(3), 535-538.

Panova, M.; Blakeslee, A.M.; Miller, A.W.; Mäkinen, T.; Ruiz, G.M.; Johannesson, K.; André, C. (2011). Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. PloS one, 6(3)

Pérez-Figueroa, A.; García-Pereira, M.; Saura, M.; Rolán-Alvarez, E.; Caballero, A. (2010) Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23(10), 2267-2276.

Rafajlović, M.; Eriksson, A. et al. (2013) The Effect of Multiple Paternity on Genetic Diversity of Small Populations during and after Colonisation. PloS one 8(10)

Raffaelli, D.; Hawkins, S. (1996) Intertidal Ecology. Kluwer Academic Publishers.

Reid, D.G. (1996) Systematics and evolution of Littorina. London, Ray Society.

Reimchen, T.E. (1979) Substrate heterogeneity, crypsis, and colour polymorphism in an intertidal snail (*Littorina mariae*). Canadian Journal of Zoology 57, 1070-1085.

Rice, W.R. (1989) Analyzing tables of statistical tests. Evolution 43, 223-225.

Roesti, M.; Hendry, A.P.; Salzburger, W.; Berner, D. (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. Molecular Ecology 21(12), 2852–286.

Rohlf, F.; Bookstein F. (2003) Computing the uniform component of shape variation. Systematic Biology 52(1), 66–69.

Rolán, E.; Templado, J. (1987) Consideraciones sobre el complejo *Littorina obtusata-mariae* (Mollusca, Gastropoda, Littorinidae) en el Noroeste de la península Ibérica. Thalassas 5, 71–85.

Rolán-Alvarez, E.; Zapata, C.; Alvarez, G. (1995) Distinct genetic subdivision in sympatric and sibling species of the genus *Littorina* (Gastropoda: Littorinidae). Heredity 75, 1-9.

Rolán-Alvarez, E. (2007) Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. Journal of Molluscan Studies 73(1), 1-10.

Rosenberg, N. (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4, 137–138

Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular ecology resources 8(1), 103-106.

Rundle, H.; Nagel, L.; Boughman, J.; Schluter, D. (2000) Natural selection and parallel speciation in sympatric sticklebacks. Science 287(5451), 306-308.

Saitou, N.; Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4(4), 406-425.

Salazar, C.; Baxter, S.W. et al. (2010). Genetic evidence for hybrid trait speciation in Heliconius butterflies. PLoS genetics 6(4)

Schluter, D. (2000) The Ecology of Adaptive Radiation. Oxford, Oxford University Press.

Schmidt, P.; Phifer-Rixey, M.; Taylor, G.; Christner, J. (2007) Genetic heterogeneity among intertidal habitats in the flat periwinkle, *Littorina obtusata*. Molecular Ecology 16, 2393–2404.

Seeley, R.B. (1986) Intense natural selection caused a rapid morphological transition in a living marine snail. Proceedings of the National Academy of Sciences 83(18), 6897-6901.

Seehausen, O.; Butlin, R. et al. (2014) Genomics and the origin of species. Nature Reviews Genetics 15, 176–192.

Sinnock, P. (1975) The Wahlund effect for the two-locus model. American Naturalist 109(969), 565-570.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139(1), 457-462.

Sokal, R.; Rohlf, F. (1994) Biometry. New York, Freeman & Company.

Tatarenkov, A.; Johannesson, K. (1994) Habitat related allozyme variation on a microgeographic scale in the marine snail *Littorina mariae* (Prosobranchia:Littorinacea) Biological Journal of the Linnean Society 53, 105-125.

Tatarenkov, A.; Johannesson, K. (1998) Evidence of a reproductive barrier between two forms of marine periwinkle *Littorina fabalis* (Gastropoda). Biological Journal of the Linnean Society 63, 349–365.

Tatarenkov, A.; Johannesson, K. (1999) Micro- and macrogeographic allozyme variation in *Littorina fabalis*; do sheltered and exposed forms hybridize? Biological Journal of the Linnean Society 67, 199-212.

Trussel, G.; Etter, R. (2001) Integrating genetic and environmental forces that shape the evolution of geographic variation in a marine snail. Genetica 112, 321–337.

Vekemans, X.; Beauwens T.; Lemaire, M.; Roldán-Ruiz, I. (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology 11, 139–151.

Vos, P.; Hogers, R. et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23(21), 4407-4414.

Wares, J.; Cunningham, C. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. Evolution 55(12), 2455–2469.

Weir, B.; Cockerham, C. (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6), 1358-1370.

Westram, A.; Galindo, J.; Rosenblad, M.; Grahame, J.; Panova, M.; Butlin, R. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? Molecular Ecology 23(18), 4603-4616.

Whitlock, R.; Hipperson, H.; Mannarelli, M.; Butlin, R.K.; Burke, T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. Molecular Ecology Resources 8, 725-735.

Wilding, C.S.; Butlin, R.; Grahame, J. (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of Evolutionary Biology 14(4), 611-619.

Williams, G.A. (1990) The comparative ecology of the flat periwinkles, *Littorina obtusata* (L.) and *L. mariae* Sacchi et Rastelli. Field Studies 7, 469-482.

Wood, H.; Grahame, J.; Humphray, S.; Rogers, J.; Butlin, R. (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. Molecular Ecology 17(13), 3123-3135.

Zardi, G.I.; Nicastro, K.R.; Ferreira Costa, J.; Serrão, E.A.; Pearson, G. A. (2013) Broad scale agreement between intertidal habitats and adaptive traits on a basis of contrasting population genetic structure. Estuarine, Coastal and Shelf Science 131, 140-148.

Zhivotovsky, A. (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology 8, 907–913.

Supplementary Information

Material and Methods

1. Prospection

Despite several descriptions of the distribution of flat periwinkles in Northern Europe (NE) and in Galicia (Rolán & Templado, 1987; Kemppainen et al., 2011; Reid, 1996), their presence in Portugal was basically unknown. In order to fill this gap, an initial prospection along the Portuguese coast from Caminha to Nazaré was carried out between 2011 and 2013. Visits to the locations were performed during the lowest tides of each month (< 0.7m), in the two hours around the diurnal low tide (one hour before and one after), to avoid biasing prospection against ecotypes inhabiting the lower part of the intertidal, which are mostly visible during the low tide. In these visits, the presence of *L. obtusata* and *L. fabalis* (with its different ecotypes) was recorded, together with the species of macroalgae where they were found. Locations were photographed and the coordinates registered using a Global Positioning System (GPS) device (Garmin Dakota 10). This information was used to select the sampling sites for further geometric-morphometrics and genetic analyses.

2. Sampling

In the Iberian Peninsula (IP), at least two locations were sampled for each *L. fabalis* ecotype while one *L. obtusata* population was sampled for comparison (Figure 8; Table 2). The size of the sampling area was a compromise between avoiding the collection of individuals from different subpopulations and, in the other extreme, sampling only related or inbred individuals. Thus, individuals were collected from areas comprising about 10 m^2, except when densities were low, in which case we increased the sampling area until a maximum of about 200 m^2. Importantly, we tried to not bias our collection towards phenotypically pure individuals from each ecotype but rather to represent all the phenotypic variation in terms of shell shape and color present in each sampling location. Individuals were brought alive to the laboratory and were frozen at -20°C.

In NE, replicate samples for each ecotype were defined at two geographic scales: local (within each country, with the distance between sampling sites for each ecotype < 100 Kms) and regional (between countries, with the distance between sampling sites for each ecotype > 1000 Kms), in order to investigate how independent is ecotype divergence in these populations of *L. fabalis* at different resolution levels (Figure 9; Table 3). In each site, sampling was performed following the same protocol as in the IP. Individuals were stored in ethanol to facilitate their transport to the laboratory and then they were frozen at 20°C.

3. Sample processing

All collected samples were processed at ECIMAT marine station (Estación de Ciencias Mariñas de Toralla, University of Vigo, Vigo, Spain). The snails were removed from their shells and they were sexed under a dissection microscope (Nikon SMZ1000). A pre-classification of individuals into *L. fabalis* or *L. obtusata* was done based on the penis morphology in the case of males, and in shell appearance in the case of females. The soft tissue was stored in ethanol until the DNA extraction was performed and the shells were photographed for geometric morphometrics analysis. As an exception, ME individuals from Silleiro and Oia were not photographed because it was not possible to remove the snail without damaging the shell and, consequently, they were only included in the genetic analysis. Because samples from NE were initially stored in ethanol, it was difficult to remove the tissues from the shell. Thus, contrary to the samples collected in the IP, the shells were first photographed with the individuals inside and then gently crushed to maintain the tissues intact for sexing and further DNA extraction.

4. Geometric Morphometrics Analysis

The morphometric analysis of a biological system involves photographing specimens in a standardized position and placing landmarks (LMs) on the photos. It provides an accurate estimate of the size and shape of each specimen, allowing an objective characterization of phenotypes for different species or ecotypes.

Landmarks are a set of morphometric coordinates that represent the location of significant biological features that can be used to describe the shape of an individual (Kaliontzopoulou, 2011). In addition, semilandmarks (SLMs), which represent sliding rather than fixed points (i.e. LMs) in a geometric surface, can also be used to capture information about curvatures. A preliminary analysis conducted in a subset of individuals from the IP revealed that 36 coordinates were necessary to fully characterize the shell shape (4 LMs + 32 SLMs). However, since the individuals from NE were still inside the shell when photographed, the coordinates from the inner aperture could not be placed (Figure 10), and thus only 28 points (2 LMs + 26 SLMs) were considered. In the IP, LM1 and LM2 were defined by the points where the aperture connects with the shell, and LM 5 and LM6 as the upper and lower points of the internal aperture, respectively. In NE, only LM1 and LM2 were used.

4.1. GM pipeline

The software packages tpsUtil (v.1.58), tpsDig (v.1.40) and tpsRelw (v.1.49) (http://life.bio.sunysb.edu/ee/rohlf/software.html) were used to perform the Generalized Procrustes Analysis, based on the superimposition method (Kaliontzopoulou, 2011). After photographing the individuals, the tpsUtil is used to create a file with all the images to be analyzed and that file is used in tpsDig, where the LMs and SMLs are positioned and a scale is set (using the graph paper included in the background of each photo). The tpsUtil is then used to obtain the landmark configuration for each individual and that file is used in tpsRelw, where all the landmark configurations are superimposed to create a consensus shape (i.e. the landmark configuration that describes the general shape trend of a population, ecotype or species). The tpsRelw allows the assessment of the centroid size (CS) for each individual and, after correcting the effects of position, rotation and translation of the shell (also done in tpsRelw), the grid of each individual is warped until its LMs and SLMs coincide with the consensus, and shape differences are extracted for each individual and visually represented by deformation grids.

4.2. Shape differences

Shape differences are subdivided into uniform and non-uniform components (Figure S1). Among the first, those that do alter shape (the first uniform component, U1 and the second, U2) describe variation that affects all the coordinates with the same intensity. U1 expresses changes in the horizontal scale, maintaining the vertical axis' coordinates fixed and allowing the horizontal ones to move (i.e. compression or dilatation), U2 only allows the vertical axis' coordinates to move (i.e.

shearing) (Rohlf & Bookstein, 2003; Zelditch et al., 2004). Non-uniform components (Relative Warp Scores, RWs) describe non-linear local deformations on the shell, and its number (n) depends on the number of LMs and SLMs that were defined, following the relation n= 2((LM+SLM)−4)) (Zelditch et al., 2004). In both cases, the percentage of variation explained by the components diminishes from the first to the last component (Rohlf & Bookstein, 2003; Kaliontzopoulou, 2011).

4.3. Data analysis

Normality tests were performed to investigate if the variables (Centroid Size - CS, Relative Warps - RW1-3, and Uniform Components - U1 and U2) conformed to a normal distribution within each *L. fabalis* ecotype and within *L. obtusata* using the Shapiro-Wilk test (Shapiro & Wilk, 1965). In most cases, the variables followed a normal distribution and so one-way ANOVA tests were performed to check for significant differences in the means across the ecotypes (independent variable) and across each ecotype and *L. obtusata*. In the IP dataset, given the peculiarities of the individuals collected in Cabo do Mundo, the morphological differences between this population and *L. obtusata* as well as *L. fabalis* were also tested.

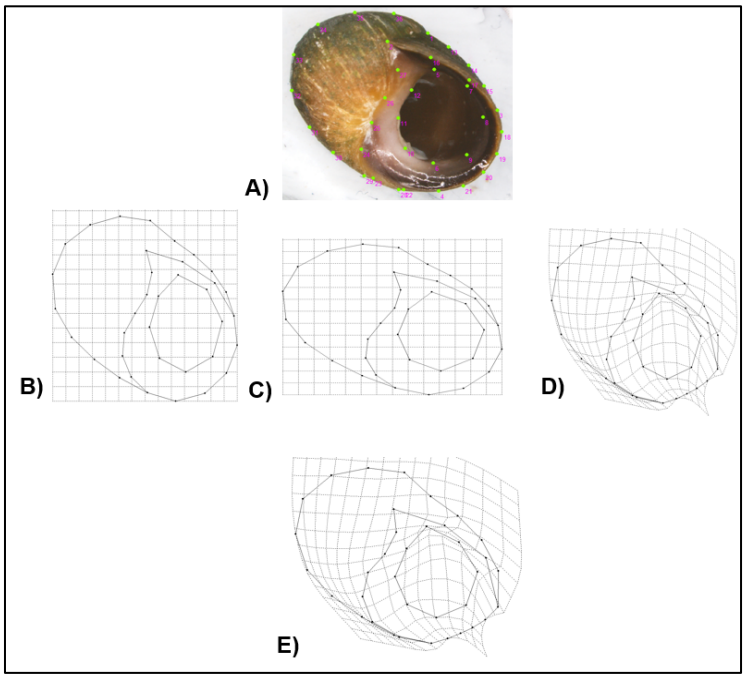

Figure S1. Deformations of body shape in a *Littorina fabalis* specimen, represented by deformation grids. A) Specimen with digitized LMs and SLMs. B) Digitized specimen from A represented by the connections between LMs and SLMs. C) Uniform components (U1 and U2). D) Non-uniform components; E) Deformation grid with all shape components

Post-hoc tests (Tukey HSD) were performed to inspect for significant differences in pairwise comparisons between the different ecotypes, as well as between these and *L. obtusata* in the IP. Since the normality of some variables was rejected in a few cases, non-parametric tests (Mann-Whitney U) (Mann & Whitney, 1947) were also performed. Additionally, in NE, a student *t*-test was

performed to compare CS in a pairwise fashion between sheltered and exposed ecotypes at each region and between sheltered or exposed ecotypes across the three regions.

A principal component analysis (PCA) was performed for each dataset. This statistical procedure converts a set of variables into linearly uncorrelated variables, called principal components (PCs), in such a way that the first principal component (PC1) accounts for the highest percentage of variability in the data, while the last accounts for the lowest. PCAs were performed with and without CS as a variable in order to investigate shape differences alone versus the effect of size differences in the overall morphological differences.

5. Genetic Analyses

5.1. Microsatellites

5.1.1. Laboratorial procedures

Microsatellite loci development was performed by GENOSCREEN (Lille, France) using high-throughput pyrosequencing (i.e. 454) of microsatellite repeat enriched libraries from a pool of nine samples of *L. fabalis* DNA (FI ecotype), following the protocol described by Malausa et al. (2011). Loci were chosen according to the following criteria: similar melting temperature for all loci, non-overlapping (predicted) size of PCR products, high GC content, representation/inclusion of di-, tri- and tetranucleotide motifs. Initially, 33 primer pairs were tested in single PCRs in one *L. fabalis* (Tirán) and one *L. obtusata* (Moinhos) population. This step was useful to confirm the expected size range of the alleles and to discard monomorphic loci or loci that failed to amplify. The single PCRs were performed using the same conditions as the multiplex PCRs described in the methodology section of this work, but using a single primer pair. A detailed list of all tested primer pairs can be found in Table S1.

In principle, primers that do not present special problems in single PCRs are expected to behave in a similar manner in multiplex reactions. However, two main problems were observed: i) some primers introduced a lot of noise when multiplexed, probably due to unspecific amplification; ii) some primers were weakly amplified in multiplex conditions, due to competition during the PCR. Sixteen of the 33 loci initially tested were discarded due to i), ii) or both (Table S2).

In the end, a total of 17 microsatellite loci (seven loci with a tetranucleotide repetition motif, five tri- and other five with a dinucleotide motif) were distributed in three multiplex reactions: multiplex 1 with six loci (PBL8, 193Q, QVOM, KJ2E, 881, VPVX), multiplex 2 with six loci (1871, ZIBW, LHYM, 927, EKYY, XENN) and multiplex 3 with five loci (EVLS, DAEH, 47, TEM7, ZR6M) (see Table S1 for details).

Table S1. Summary of all the tested microsatellite loci. Size refers to the predicted size obtained with the 454 pyrosequencing. Tm F and Tm R are the melting temperature of the forward and reverse primer, respectively. Motif indicates the sequence of nucleotides that is repeated and N rpts indicates the number of times the motif is repeated. Primers marked with [1] were included in multiplex 1; those marked with [2], in multiplex 2; and those marked with a [3], in multiplex 3.

Name	Dye	Size	TmF	TmR	Forward primer (5'-3')	Reverse primer (5'-3')	Motif	N rpts
[1]PBL8	HEX	197	60	62	CCCAGACAATGCAGCCTAC	CGGTAACTGAGTTGTGCAGC	gttt	12
[2]1871	HEX	105	60	60	CACCCACCCCTATTACCCA	GGGTTGATGGATGAGTGGAT	atcc	5
FBV4	FAM	283	62	60	CTTAGAGCCAAAGCAGCACC	CAACGACGTATGTGCAAGGA	ctgt	5
SSMD	FAM	129	60	60	TGCGATGCTAACTTTTGTCTG	CTTGAATGTGCCAGGGTTTC	gtgc	5
0ZZ9	FAM	152	60	62	CCATCTCACACGGCATATTG	CCTCCTCCACCGTTACAATC	tgga	11
0DPQ	HEX	293	62	58	CAGACCGTCGCGATATAACC	AGCTCCGTTTCAATCTCCAA	aaag	5
[1]193Q	FAM	215	62	58	TTTGCATACACCCGTCTAACC	GCTATTTCATTAAGCCGCCA	caaa	9
CLEU	FAM	102	58	58	TGAAAACGACGTTAAACACCA	TTCTTCGGAACGCTGAAAAC	aaag	5
11EZ	HEX	140	62	62	GAAGAAACTGACGAACATTTGC	TGTACGTGACTGTCTGTCGG	agac	8
[1]QVOM	NED	117	62	62	ACATGGGATACGACTACCCG	AGCCTAGCTGCTACGTCCAA	aaac	10
DSQG	NED	241	58	60	TCTGAATAAATTCCGAAAATGG	TCGAAGTGTCAGAGGTTTGC	tctg	9
[3]EVLS	NED	112	58	62	GTTTTGGTTGAATGTTGGGC	GACAGAAAACAGAAACAACGAAA	agtc	5
[3]DAEH	NED	242	60	60	ACCGCACAGCTACACGAAG	TCGTGTTTCATGATGCCCTAT	tgtt	5
AIU6	HEX	90	60	62	AAGTGTAGCCTATGCGATGC	ATCGATAGACTCGGAAATGTAAA	gt	7
K94Y	NED	108	60	60	CTGGGCGTTAAGCAAACAAG	GCATCTGCTGAAGGGACATT	tggt	10
M82S	FAM	166	60	60	CATATCAGGGCGGGTTTAAG	CTGATACTGGCCCCTTCGT	ttgt	5
[3]47	HEX	194	62	62	TGTTGCTCTGCAGATTATGACA	GATCGATGCCCTGACATAGC	tc	8
9U8S	NED	212	62	60	ACTGGGATGTCAACGTAGGG	GAACCTCGTCATCTTTTGGC	ct	5
[3]TEM7	FAM	237	60	60	CTCATGCTGTTCCTGGTTGA	TGCGTGGTTTAAATTGTTCTTG	ac	5
[3]ZR6M	FAM	105	60	62	TGAGACATGAAGCCTGTGCT	AATACAATCTGGTGTCTGCGG	aaac	6
1222	HEX	119	60	62	TCTTGACTCGACGAGGTGG	CCTGCAAACCCTAACACATTC	tgt	5
537	NED	140	62	58	CATCGTGGAGAATACCTGGG	TGGCAAACACAGAAACAAACA	gttt	7
[1]KJ2E	HEX	245	62	60	TCACTTACCTCAAACCTTGCG	CCACAGGCGGGGTGTAAG	gct	5
1027	NED	291	60	60	GGTATCTTTTCTTGAGCCCG	TGTATCTTCGTGTGCTGGGA	gtt	5
[1]881	FAM	316	58	62	ACGCCCAGAATTGCCTAAAT	GCTTGTTTATTGACAGGCAGC	gtt	22
[2]ZIBW	NED	96	58	58	TTTTGTTAACACGTGGCAGTT	TTGGTGAGTGCGTGCATTAT	ca	11
[2]LHYM	FAM	192	62	58	TGGTACGGACGAGGCTCTTA	ATTGCTTGAATGCCCGTTAC	ac	12
[2]927	HEX	241	62	62	CATACAATCCGTCCCTCTCC	TACTCGAACAGGAACGAGGC	ag	9
D8DU	FAM	91	60	60	ACCCGTAGCGAACACTGAAA	CACTTTAACGCAGAACGCAG	ctt	7
[2]EKYY	HEX	145	60	60	TTGTCAAGAATGTTGGTTCCC	ATCCGGAATCGACAAGTGAC	ctt	8
[2]XENN	NED	242	58	58	CAGCACAAGGCGGTTCAG	TCCTATTTGAAGATGCGGTG	caa	10
1173	HEX	116	60	62	CACGACAATCCAACAACACC	TTGACTGAGAAAGAAAGAAAACG	ac	14
[1]VPVX	NED	198	58	58	CGCTACGCCACTTCGTTTA	AATCGGAGAACAAAACCACG	ttg	17

Table S2. List of discarded microsatellite loci. The reason presented for discarding each locus does not reflect all of the problems associated with that primer pair and is only intended to be a very general explanation of the encountered problems.

Locus	Reason to be discarded
FBV4	Noise in multiplex reactions
SSMD	No amplification in many samples
0ZZ9	Noise in multiplex reactions
0DPQ	Monomorphic
CLEU	Low variation
11EZ	Extra peaks
DSQG	Low variation
AIU6	Low peaks
K94Y	Noise and low peaks
M82S	No amplification
9U8S	Low variation
1222	Monomorphic
537	Monomorphic
1027	Noise and low variation
D8DU	Noise in multiplex reactions
1173	No amplification

5.2. AFLPs

5.2.1. Laboratorial procedures

Initially, 100 ng of total genomic DNA were digested in a final reaction volume of 12 µL with 4U *Eco*RI (New England Biolabs, NEB) and 2U *Mse*I (NEB) in 1X Buffer EcoRI (NEB) supplemented with 0.03 µg of BSA (bovin serum albumin) for 3.5 h at 37°C. Enzymes were then inactivated by heat shock at 70°C during 10 min. All the samples were randomly distributed across the 96-well reaction plates used in this study and each plate included replicates of individuals present in other plates. A total of 15% of individuals were repeated. This design was maintained along the following AFLP steps.

Ligation reaction was performed by adding, to the digestion reaction, 3 µL of a solution containing 5 pmol of EcoRI adapter, 5 pmol of MseI adapter, 0.25 U T4 DNA ligase (Roche) in 1X Ligase Buffer. The samples were then incubated for 16 h at 16°C.

Pre-selective PCR was performed in 10 µL final volume containing 2 µL of diluted ligation (1:4 dilution), 0.3 mM of dNTP mix, 2 mM of $MgCl_2$, 5 pmol of Eco+A primer, 5 pmol of Mse+C primer, 0.3 U of *Taq* polymerase (Bioline) in 1X PCR Buffer. PCR conditions consisted of an initial step at 72°C for 2 min, followed by 20 cycles of 94°C for 20 s, 56°C for 30 s and 72°C for 2 min, and a final step at 72°C for 10 min.

Selective PCRs were performed on 1 µL of diluted pre-selective PCR (1:4 dilution) using the same conditions as for the pre-selective PCR but with the addition of 4 pmol of Eco+ACT (FAM labeled), 2.5 pmol of Eco+AAG (NED labeled) and 5 pmol of Mse+3 primers in each reaction. Two different selective PCRs were performed, one with Mse+CAA and the other with Mse+CAC. PCR conditions started with a denaturing step at 94°C for 2 min, followed by 10 cycles of 94°C for 20 s, 66°C (decreasing 1°C at every cycle) for 30 s and 72°C for 2 min, then followed by 20 cycles of 94°C for 20 s, 56°C for 30 s and 72°C for 2 min, and a final extension step at 72°C for 10 min. See Table S3 for adapter and primer sequences used in the AFLP protocol.

Table S3. Nucleotide sequences for adapters, pre-selective and selective PCR primers used in the AFLP protocol. Selective nucleotides in each primer are highlighted in bold. FAM and NED are the fluorochromes used to label the primers.

Adapters	
Eco adaptor	5'-CTCGTAGACTGCGTACC-3' 3'-CATCTGACGCATGGTTAA-5'
Mse adaptor	5'-GACGATGAGTCCTGAG-3' 3'-TACTCAGGACTCAT-5'
Pre-selective primers	
Eco + A	5'-GACTGCGTACCAATTC **A**-3'
Mse + C	5'-GATGAGTCCTGAGTAA **C**-3'
Selective Primers	
Eco + ACT (FAM)	5'-GACTGCGTACCAATTC **ACT**-3'
Eco + AAG (NED)	5'-GACTGCGTACCAATTC **AAG**-3'
Mse + CAA	5'-GATGAGTCCTGAGTAA **CAA**-3'
Mse + CAC	5'-GATGAGTCCTGAGTAA **CAC**-3'

Results

1. Geometric Morphometrics

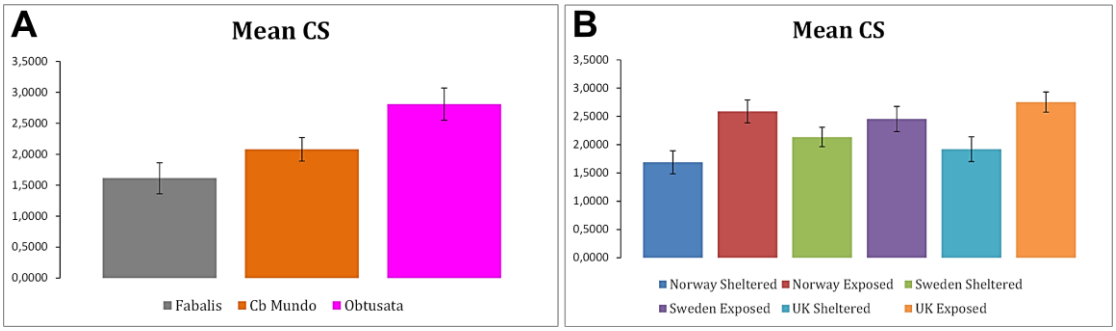

Figure S2. A) Mean centroid size (CS) and standard deviation for *L. fabalis*, Cabo do Mundo and *L. obtusata* in the IP. Mean CS is significantly different between the three presented groups (p<0.0001). B) Mean CS and standard deviation for each region and ecotype in NE. Means CS is significantly different between all groups (p<0.01), except between Norway Exposed and Sweden Exposed.

Table S4. One-way ANOVA using data for CS, RW1-3, U1 and U2 from *L. obtusata* and the three *L. fabalis* ecotypes in the IP.

Effect	Test	Value	F	Effect df	Error df	*p-value*
Intercept	Wilks	0.0193	1480.716	6	175	0.000*
Ecotype	Wilks	0.1264	29.658	18	495.46	0.000*

* indicates significant values

Table S5. t-Test comparing the centroid size between all the Iberian populations.

	Comparison	*t-value*	*p-value*
Within FI			
	Canido – Cangas	-8.7804	0.0000*
	Canido - Muros	-3.7733	0.0007*
	Canido – Abelleira	-3.5414	0.0012*
	Canido – Tirán	-8.0790	0.0000*
	Cangas – Muros	2.9746	0.0068*
	Cangas – Abelleira	3.8178	0.0007*
	Cangas – Tirán	2.3806	0.0219*
	Muros - Abelleira	0.4297	0.6712
	Muros – Tirán	-1.7826	0.0824
	Abelleira - Tirán	-2.6518	0.0112*
Within ME	Agudela - Póvoa	-0.9421	0.3537
Within ZS	Grove 1 – Grove 2	2.1108	0.0470*
Within *L. obtusata*	Cb. do Mundo - Moinhos	-9.0413	0.0000*
Between FI - ME			
	Canido - Póvoa	4.3929	0.0001*
	Canido - Agudela	5.3646	0.0000*
	Agudela - Cangas	-12.6571	0.0000*
	Agudela – Muros	-7.2357	0.0000*
	Agudela – Abelleira	-7.3833	0.0000*
	Agudela – Tirán	-13.0220	0.0000*
	Póvoa – Cangas	-11.2723	0.0000*
	Póvoa – Muros	-6.2062	0.0000*
	Póvoa – Abelleira	-6.1377	0.0000*
	Póvoa – Tirán	-10.7964	0.0000*
Between FI - ZS			
	Canido – Grove 1	-3.5850	0.0011*
	Canido – Grove 2	-1.2104	0.2362
	Cangas – Grove 1	5.3689	0.0000*
	Cangas – Grove 2	6.1177	0.0000*
	Grove 1 – Muros	-1.1305	0.2699
	Grove 1 – Abelleira	-0.6640	0.5123
	Grove 1 – Tirán	-3.9113	0.0003*
	Grove 2 – Muros	-2.2283	0.0389*
	Grove 2 – Abelleira	-1.9238	0.0674
	Grove 2 – Tirán	-5.0502	0.0000*
Between ME -ZS			
	Agudela – Grove 1	-8.5810	0.0000*
	Agudela – Grove 2	-5.7461	0.0000*
	Póvoa – Grove 1	-8.6449	0.0000*
	Póvoa – Grove 2	-6.7314	0.0000*
Between FI – *L. obtusata*			
	Cb. do Mundo - Cangas	2.1954	0.0395*
	Cb. do Mundo - Muros	4.4109	0.0003*
	Cb. do Mundo – Abelleira	5.3347	0.0000*
	Cb. do Mundo – Tirán	4.8138	0.0000*
	Canido - Moinhos	-24.8726	0.0000*
	Canido – Cb. do Mundo	-10.4047	0.0000*
	Cangas - Moinhos	-13.4884	0.0000*
	Abelleira - Moinhos	-17.2870	0.0000*
	Tirán- Moinhos	-21.0419	0.0000*
Between ME – *L. obtusata*			
	Cb. do Mundo – Póvoa	12.7302	0.0000*
	Cb. do Mundo – Agudela	13.6718	0.0000*
	Agudela - Moinhos	-27.8966	0.0000*
	Póvoa - Moinhos	-22.6002	0.0000*
Between ZS – *L. obtusata*			
	Cb. do Mundo – Grove 1	7.3418	0.0000*
	Cb. do Mundo – Grove 2	7.7499	0.0000*
	Grove 1 – Moinhos	-18.9568	0.0000*
	Grove 2 - Moinhos	-16.7349	0.0000*

* indicates significant values

Table S6. Tukey HSD between each Iberian *L. fabalis* ecotype for all the variables. Numbers in brackets indicate the number of individuals analyzed. Values in bold at the diagonal indicate the mean value for the corresponding ecotype.

	CS			**RW1**		
Ecotype	FI (92)	ME (32)	ZS (23)	FI (92)	ME (32)	ZS (23)
FI	**1.721**			**0.005**		
ME	0.000*	**1.312**		0.636	**-0.008**	
ZS	0.014*	0.000*	**1.597**	0.711	1.000	**-0.008**
	RW2			**RW3**		
Ecotype	FI (92)	ME (32)	ZS (23)	FI (92)	ME (32)	ZS (23)
FI	**0.009**			**0.004**		
ME	0.097	**-0.005**		0.621	**-0.001**	
ZS	0.000*	0.018*	**-0.030**	0.015*	0.214	**-0.016**
	U1			**U2**		
Ecotype	FI (92)	ME (32)	ZS (23)	FI (92)	ME (32)	ZS (23)
FI	**-0.001**			**0.000**		
ME	0.396	**0.008**		0.015*	**0.010**	
ZS	0.708	0.224	**-0.008**	0.008*	0.000*	**-0.013**

* indicates significant values

Table S7. One-way ANOVA using data for CS, RW1-3, U1 and U2 from the two *L. fabalis* ecotypes in NE.

Effect	**Test**	**Value**	**F**	**Effect df**	**Error df**	***p***
Intercept	Wilks	0.0110	1557.7070	4.000	71.000	0.0000*
Sex	Wilks	0.9310	1.3210	4.000	71.000	0.2710
Ecotype	Wilks	0.2320	58.5990	4.000	71.000	0.0000*
Sex*Ecotype	Wilks	0.9940	0.1130	4.000	71.000	0.9780

* indicates significant values

Table S8. t-Test comparing ecotypes across the three countries in NE.

CS	**Norway - Sweden**	**Norway - UK**	**Sweden – UK**
Sheltered ecotype	t=-5.4804 p=0.0000*	t=-2.5078 p=0.0197*	t=2.9821 p=0.0057*
Exposed Ecotype	t=1.2776 p=0.2168	t=-2.5441 p=0.0160*	t=-3.1817 p=0.0045*
RW1			
Sheltered ecotype	t=1.5027 p=0.1485	t=0.5928 p=0.5591	t=-1.1834 p=0.2462
Exposed Ecotype	t=0.4755 p=0.6398	t=-1.0528 p=0.3003	t=-1.2947 p=0.2095
RW3			
Sheltered ecotype	t=2.5749 p=0.0181*	t=2.9200 p=0.0077*	t=0.9504 p=0.3498
Exposed Ecotype	t=-2.0167 p=0.0581	t=-3.2724 p=0.0026*	t=0.3807 p=0.7073
U1			
Sheltered ecotype	t=1.6575 p=0.1130	t=0.1935 p=0.8482	t=-1.9189 p=0.0649
Exposed Ecotype	t=-0.5904 p=0.5619	t=-0.8611 p=0.3956	t=0.0397 p=0.9687

* indicates significant values

2. Microsatellites

Table S9. Allele frequencies for the microsatellite loci analyzed in the Iberian populations of flat periwinkles. N is the number of individuals analyzed.

			AGU	CAG	CAN	CMU	GR1	GR2	MUR	OIA	POV	SIL	ABE	MOI	TIR
Locus	Alelles														
QVOM		N	32	20	24	24	23	19	24	22	20*	24	22**	35	35
	82		-	-	-	-	-	-	-	-	-	-	-	0.04	-
	90		-	-	-	0.54	-	-	-	-	-	-	-	0.29	-
	98		-	-	-	0.06	-	-	-	-	-	-	0.02	0.19	-
	106		-	0.03	-	0.10	0.02	-	-	0.02	-	-	-	0.34	-
	110		0.81	0.63	0.50	0.06	0.83	0.71	0.52	0.84	0.50	0.65	0.41	-	0.64
	114		-	0.08	0.06	0.06	-	-	-	0.02	-	0.04	-	0.04	0.04
	118		-	0.18	0.42	-	-	0.08	0.06	0.02	-	-	0.14	-	0.29
	122		-	-	-	-	-	-	0.02	-	-	-	-	0.09	-
	126		-	-	-	-	-	0.03	0.02	-	0.10	-	0.09	0.01	-
	130		-	-	-	-	-	-	0.10	0.02	-	-	0.02	-	-
	134		-	-	-	-	-	-	-	-	-	-	0.02	-	-
	138		0.19	0.10	0.02	0.17	0.15	0.18	0.27	0.07	0.40	0.27	0.30	-	0.03
	170		-	-	-	-	-	-	-	-	-	0.04	-	-	-
VPVX		N	32	20*	24*	24	22	20	24	22	22	24	23	35	35
	176		-	-	0.08	-	-	0.03	-	-	-	0.04	-	0.01	0.01
	180		0.72	0.05	0.06	0.17	0.09	0.20	0.10	0.05	0.57	0.33	-	0.51	0.07
	183		-	-	-	-	0.02	0.05	0.19	-	-	-	0.07	-	-
	186		-	0.03	-	0.60	0.07	-	0.08	0.07	-	-	0.15	0.10	0.01
	189		0.09	-	0.13	0.06	0.20	0.43	0.27	0.18	0.09	0.58	0.15	-	-
	190		-	-	-	-	-	-	-	-	-	-	-	0.21	0.09
	192		-	-	0.08	-	0.05	0.03	0.04	0.14	-	0.02	0.09	0.16	0.06
	195		-	-	-	-	-	-	-	-	-	-	-	-	0.07
	196		0.05	0.13	0.04	0.06	0.09	0.10	0.17	0.02	0.34	0.02	0.07	-	-
	199		-	-	-	-	-	0.08	0.06	-	-	-	0.09	-	-
	202		-	0.03	-	-	0.09	0.03	0.02	0.02	-	-	0.20	-	-
	205		-	0.05	0.08	0.08	-	0.03	0.06	0.02	-	-	0.09	-	0.04
	208		0.06	-	-	0.02	0.11	-	-	-	-	-	0.04	-	0.06
	211		-	0.10	0.06	-	-	-	-	-	-	-	0.02	-	0.11
	214		0.08	0.03	-	-	0.05	-	-	0.05	-	-	-	-	0.13
	218		-	0.08	0.02	-	0.07	-	-	0.09	-	-	-	-	0.11
	221		-	0.25	0.23	-	0.02	0.05	-	0.09	-	-	0.02	-	0.10
	224		-	-	-	-	0.11	-	-	0.07	-	-	0.02	-	0.07
	227		-	0.03	0.04	-	-	-	-	0.02	-	-	-	-	-
	230		-	0.20	0.04	-	-	-	-	0.05	-	-	-	-	0.03
	233		-	0.03	0.06	-	-	-	-	0.02	-	-	-	-	0.03
	236		-	0.03	0.04	-	-	-	-	0.02	-	-	-	-	-
	240		-	-	0.02	-	-	-	-	-	-	-	-	-	-
	249		-	-	-	-	-	-	-	0.05	-	-	-	-	-
	255		-	-	-	-	0.02	-	-	0.05	-	-	-	-	-
PBL8		N	32	20	24	23	23	20	24	22	23	24	23	35	35
	173		-	-	-	-	-	-	0.02	0.02	-	-	0.02	-	-
	177		-	-	-	0.17	-	-	0.02	-	-	0.02	-	-	0.01
	181		-	0.23	0.19	0.39	0.37	0.15	0.17	0.36	0.09	0.02	0.41	0.99	0.26
	185		0.30	0.25	0.42	0.20	0.24	0.30	0.40	0.30	0.76	0.08	0.17	0.01	0.20
	189		-	-	-	0.09	0.02	-	-	-	-	-	0.13	-	-
	194		-	-	-	0.07	-	0.15	0.04	-	-	0.58	-	-	-
	198		0.66	0.13	0.06	0.09	0.04	0.23	0.13	0.05	0.15	0.08	0.11	-	0.17
	203		-	-	-	-	-	0.03	0.19	-	-	0.04	0.09	-	-
	207		-	-	0.08	-	-	-	-	0.02	-	0.06	-	-	0.04
	211		0.05	0.03	0.08	-	0.04	0.03	0.02	0.11	-	0.06	0.04	-	0.09
	215		-	0.30	0.15	-	0.24	0.08	-	0.11	-	0.04	0.02	-	0.20
	219		-	0.08	0.02	-	0.02	0.05	0.02	0.02	-	-	-	-	0.03
	223		-	-	-	-	0.02	-	-	-	-	-	-	-	-
193Q		N	14	17*	21	24*	23	14*	19*	14*	20	5	23*	35	30
	200		-	0.03	0.02	-	-	-	0.05	-	-	-	0.04	-	-
	204		-	0.03	-	0.88	-	0.07	0.08	0.07	-	-	-	0.84	-
	208		0.96	0.79	0.95	0.13	1.00	0.89	0.84	0.93	1.00	1.00	0.96	0.16	1.00
	212		0.04	0.15	0.02	-	-	0.04	0.03	-	-	-	-	-	-

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Table S9. Allele frequencies for the microsatellite loci analyzed in the Iberian populations of flat periwinkles. N is the number of individuals analyzed.

			AGU	CAG	CAN	CMU	GR1	GR2	MUR	OIA	POV	SIL	ABE	MOI	TIR
Locus	Alelles														
KJ2E		N	32	20	24	24	23	20	24	22	23	24	23	35	35*
	237		0.72	0.45	0.23	0.71	0.11	0.63	0.52	0.05	0.28	0.69	0.52	0.57	0.49
	243		0.28	0.55	0.77	0.29	0.87	0.38	0.48	0.93	0.72	0.31	0.48	0.43	0.51
	252		-	-	-	-	0.02	-	-	0.02	-	-	-	-	-
881		N	32	20	24	24	23	20	24	22*	23	24	23	35	35
	281		-	-	0.02	-	-	-	-	-	-	-	-	-	0.07
	284		-	-	-	-	-	-	-	-	-	-	-	0.03	-
	287		-	-	-	0.33	-	-	-	-	-	-	-	-	-
	290		-	-	-	0.04	-	-	-	-	-	-	-	0.44	-
	292		-	-	-	0.27	-	0.03	-	-	-	-	-	-	-
	293		0.11	-	-	0.10	-	-	-	-	0.02	0.02	-	0.23	-
	296		0.06	-	-	-	-	0.05	-	0.05	-	-	0.09	0.04	-
	298		-	0.03	-	-	-	0.18	-	0.02	-	-	0.09	-	-
	299		-	0.05	-	0.10	0.02	0.08	-	0.05	-	-	0.09	-	-
	303		-	0.15	0.10	0.06	0.09	0.10	0.10	-	-	0.23	0.02	-	-
	305		-	0.20	-	0.08	-	0.08	-	0.02	-	-	0.02	-	-
	306		0.36	0.10	-	-	0.02	0.10	0.08	0.32	0.78	-	0.13	-	0.01
	309		0.13	0.13	0.04	-	0.02	0.15	0.08	0.14	-	0.33	0.13	0.03	0.14
	312		0.30	0.15	0.21	-	0.07	0.10	0.04	0.16	-	0.02	0.07	0.09	0.14
	315		0.05	0.18	0.08	-	0.26	-	0.21	0.18	-	0.13	0.11	0.07	0.19
	319		-	-	0.31	-	0.09	0.08	0.17	0.07	-	0.17	0.04	-	0.23
	322		-	0.03	0.08	-	0.13	0.03	0.04	-	-	0.04	-	-	0.11
	325		-	-	0.08	-	0.20	-	0.06	-	0.20	0.04	0.02	0.07	0.07
	328		-	-	0.04	-	0.11	-	0.04	-	-	0.02	0.04	-	0.01
	331		-	-	0.02	-	-	0.05	0.02	-	-	-	0.04	-	0.01
	334		-	-	-	-	-	-	0.08	-	-	-	0.02	-	-
	338		-	-	-	-	-	-	0.02	-	-	-	0.07	-	-
	341		-	-	-	-	-	-	0.04	-	-	-	0.02	-	-
	344		-	-	-	-	-	-	-	-	-	-	-	-	-
	347		-	-	-	-	-	-	-	-	-	-	-	-	-
	356		-	-	-	0.33	-	-	-	-	-	-	-	-	-
ZIBW		N	32	20	24	24	23	20	24	22	23	24	23	35	35
	73		-	-	-	-	-	-	0.04	-	-	-	0.04	-	-
	75		-	0.30	0.17	-	0.30	0.38	0.31	0.05	-	0.06	0.13	-	0.13
	79		-	-	-	0.75	0.07	0.05	-	-	-	-	-	0.94	0.01
	81		-	-	0.02	-	0.09	0.05	0.06	-	-	-	0.26	-	-
	85		-	-	0.08	-	-	-	-	0.02	-	0.02	0.09	-	0.01
	89		1.00	0.43	0.67	0.25	0.46	0.38	0.17	0.73	1.00	0.88	0.37	0.06	0.61
	91		-	0.20	0.02	-	0.07	0.05	0.33	0.20	-	0.04	0.02	-	0.06
	95		-	0.05	0.04	-	0.02	0.10	0.08	-	-	-	0.02	-	0.13
	99		-	0.03	-	-	-	-	-	-	-	-	0.04	-	0.04
	101		-	-	-	-	-	-	-	-	-	-	0.02	-	-
1871		N	32	20	24	24	23	20	24	22	23	24	22	35	35
	121		-	-	0.08	-	-	-	-	-	-	-	-	-	-
	132		1.00	0.98	0.85	0.25	0.39	0.43	0.73	0.80	1.00	0.71	0.93	-	0.99
	133		-	-	-	-	-	-	-	-	-	-	-	0.23	-
	136		-	-	-	0.75	-	-	-	-	-	-	-	0.77	-
	140		-	-	-	-	-	-	0.02	-	-	-	0.05	-	-
	144		-	0.03	0.06	-	0.61	0.58	0.25	0.20	-	0.29	0.02	-	0.01
ZR6M		N	32	20	24	24	23	20	24	22	23	24	23	33	35
	101		0.41	0.63	0.44	0.04	0.83	0.80	0.50	-	0.26	0.02	0.67	-	0.56
	109		-	0.08	0.25	-	0.17	0.20	0.13	-	-	-	0.11	-	0.19
	117		0.08	-	-	0.02	-	-	-	0.20	-	0.19	-	-	-
	121		0.34	-	0.06	0.08	-	-	-	0.09	0.70	0.10	-	-	-
	125		-	0.05	-	-	-	-	-	0.05	-	0.04	-	-	0.01
	129		0.05	0.25	0.25	0.06	-	-	0.38	0.66	0.04	0.65	0.15	-	0.20
	133		0.13	-	-	0.06	-	-	-	-	-	-	0.07	-	0.04
	161		-	-	-	0.73	-	-	-	-	-	-	-	1.00	-

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Table S9. Allele frequencies for the microsatellite loci analyzed in the Iberian populations of flat periwinkles. N is the number of individuals analyzed.

			AGU	CAG	CAN	CMU	GR1	GR2	MUR	OIA	POV	SIL	ABE	MOI	TIR
Locus	Alelles														
LHYM		N	32*	20	24	24*	23	20	24	22	23	24	23	35	35
	182		-	-	-	-	-	-	-	-	-	-	-	0.13	-
	184		0.42	0.10	0.02	-	0.07	0.03	0.33	0.25	-	0.31	0.07	0.13	0.16
	190		-	-	-	-	-	-	-	0.02	-	-	-	-	0.03
	191		-	0.23	0.08	-	0.02	-	-	-	-	-	-	-	0.19
	192		-	0.03	0.06	0.71	0.17	0.23	0.06	0.11	-	0.19	-	0.73	0.03
	193		-	-	-	-	-	-	-	-	-	-	-	-	0.03
	194		-	0.03	-	-	0.02	0.05	-	0.02	-	0.02	-	0.01	-
	196		0.08	0.20	0.08	-	0.22	0.33	0.25	0.25	0.02	0.17	0.35	-	0.09
	198		0.48	-	0.25	0.17	0.11	-	0.06	0.16	0.98	0.13	0.07	-	0.14
	200		-	0.13	0.04	0.13	0.04	0.08	-	0.02	-	-	-	-	0.10
	202		-	0.05	0.17	-	-	-	0.13	-	-	-	0.04	-	0.04
	204		-	-	-	-	-	-	0.02	-	-	-	0.20	-	-
	206		-	-	-	-	0.04	0.13	-	0.02	-	-	0.15	-	-
	208		-	0.03	0.10	-	0.09	0.05	0.04	-	-	0.13	-	-	0.01
	210		-	-	0.08	-	0.15	0.03	-	0.11	-	-	-	-	0.01
	212		-	-	-	-	-	0.08	0.02	0.02	-	0.02	0.02	-	0.01
	213		-	0.05	-	-	0.04	-	-	-	-	-	0.07	-	-
	215		-	-	0.02	-	-	-	-	-	-	-	-	-	-
	216		-	-	-	-	-	-	-	-	-	-	-	-	0.01
	217		-	0.05	-	-	0.02	0.03	-	-	-	-	-	-	-
	218		-	-	-	-	-	-	-	-	-	-	-	-	0.01
	219		-	0.03	0.08	-	-	-	-	-	-	-	-	-	-
	220		-	-	-	-	-	-	-	-	-	-	-	-	0.04
	221		-	0.03	-	-	-	-	-	-	-	-	-	-	-
	222		-	-	-	-	-	-	-	-	-	-	-	-	0.06
	223		0.02	0.08	-	-	-	-	-	-	-	-	0.04	-	-
	224		-	-	-	-	-	-	-	-	-	-	-	-	0.03
	225		-	-	-	-	-	-	0.04	-	-	-	-	-	-
	231		-	-	-	-	-	-	0.04	-	-	-	-	-	-
	233		-	-	-	-	-	-	-	-	-	0.04	-	-	-
XENN		N	29	20	24	23*	23	19	19*	22	17*	24	22*	35	35*
	231		-	-	-	-	0.07	0.08	0.16	-	-	-	0.30	-	0.03
	234		-	-	-	-	0.09	0.03	-	-	-	-	0.20	-	-
	239		-	-	-	-	-	-	-	-	-	-	-	0.03	-
	240		0.84	0.70	0.73	0.04	0.72	0.84	0.50	0.20	0.29	0.06	0.39	-	0.74
	242		-	-	-	-	-	-	-	-	-	-	-	0.31	-
	243		-	0.03	-	0.59	0.09	-	-	-	-	-	-	0.17	0.01
	245		-	-	-	0.07	-	-	-	-	-	-	-	0.19	-
	246		-	-	-	0.02	-	-	-	-	-	-	-	-	0.01
	247		-	-	-	0.13	-	-	-	-	-	-	-	0.09	-
	248		-	-	-	0.04	-	-	0.03	0.25	-	0.40	-	0.21	-
	252		-	0.25	0.04	0.11	-	0.05	0.05	0.20	0.71	0.21	-	-	0.14
	255		0.16	-	0.19	-	-	-	0.05	0.07	-	0.06	0.02	-	0.01
	258		-	0.03	0.02	-	0.04	-	-	0.05	-	0.19	-	-	-
	261		-	-	-	-	-	-	0.08	0.23	-	0.06	0.02	-	-
	264		-	-	0.02	-	-	-	0.05	-	-	-	-	-	-
	267		-	-	-	-	-	-	-	-	-	-	-	-	0.04
	273		-	-	-	-	-	-	0.08	-	-	-	-	-	-
	276		-	-	-	-	-	-	-	-	-	-	0.07	-	-
	290		-	-	-	-	-	-	-	-	-	0.02	-	-	-
47		N													
			32	20	24	24*	23	20	24	22	23	24	23	35	35
	193		-	-	-	-	-	-	-	0.02	-	-	-	-	-
	195		0.97	0.98	1.00	0.21	1.00	1.00	0.90	0.95	1.00	1.00	1.00	-	1.00
	197		0.03	0.03	-	0.40	-	-	0.10	0.02	-	-	-	0.44	-
	199		-	-	-	0.10	-	-	-	-	-	-	-	0.40	-
	201		-	-	-	0.29	-	-	-	-	-	-	-	0.16	-
	203		-	-	-	-	-	-	-	0.02	-	-	-	-	-
TEM7		N	32	20*	24	24	23	20	24	22	23	24	23	34	35
	233		-	0.05	-	0.33	-		-	-	-	-	-	0.91	-
	235		1.00	0.95	1.00	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.09	1.00

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Table S9. Allele frequencies for the microsatellite loci analyzed in the Iberian populations of flat periwinkles. N is the number of individuals analyzed.

			AGU	CAG	CAN	CMU	GR1	GR2	MUR	OIA	POV	SIL	ABE	MOI	TIR
Locus	Alelles														
927		N	32	20	24	24*	23	20	24	22	23	24	24	34	35
	236		-	-	-	-	0.04	0.05	-	-	-	-	-	-	-
	238		0.33	0.38	0.29	0.13	0.24	0.18	0.56	0.02	0.17	0.06	0.72	-	0.26
	240		-	-	-	-	0.13	0.03	0.06	-	-	-	0.04	-	-
	242		0.42	0.25	0.48	0.02	0.46	0.53	0.33	0.23	0.72	0.27	0.13	-	0.44
	244		-	-	0.02	0.35	0.04	0.03	-	0.05	-	0.10	0.02	-	0.01
	246		-	-	-	-	0.09	0.05	-	0.05	-	-	0.07	0.26	-
	248		-	0.30	0.13	-	-	-	-	0.45	-	0.48	-	-	0.16
	250		-	-	0.02	-	-	-	-	0.05	-	0.06	-	0.06	0.01
	252		-	-	-	-	-	-	-	-	-	-	-	0.03	-
	254		-	-	-	-	-	-	-	-	-	-	-	0.03	-
	256		-	-	-	-	-	-	-	-	-	-	-	0.12	-
	258		-	-	-	0.31	-	-	-	-	-	-	-	-	-
	260		-	-	-	0.13	-	-	-	-	-	-	-	0.13	-
	262		-	-	-	0.02	-	-	-	-	-	-	-	0.18	-
	264		-	0.08	-	0.04	-	0.03	-	0.16	-	-	0.02	0.03	-
	266		0.25	-	0.02	-	-	0.10	0.04	-	0.11	0.02	-	0.01	0.11
	268		-	-	0.04	-	-	0.03	-	-	-	-	-	0.15	-
	270		-	-	-	-	-	0.05	-	-	-	-	-	-	-
	276		-	0.38	-	0.13	-	0.18	-	0.02	-	-	0.72	-	-
EVLS		N	32	20	23	24	23	20**	23*	22	23	24	23	35	35
	103		-	-	-	-	0.35	0.28	0.33	0.36	-	0.21	0.11	-	0.10
	111		1.00	0.93	0.93	1.00	0.61	0.70	0.30	0.50	0.98	0.56	0.41	1.00	0.73
	115		-	0.08	-	-	-	-	0.22	0.14	0.02	0.23	0.26	-	0.14
	120		-	-	0.07	-	0.04	0.03	0.07	-	-	-	0.11	-	0.03
	128		-	-	-	-	-	-	0.09	-	-	-	0.11	-	-
DAEH		N	32	20	23	9*	23	20	21*	21	23	24	23	0	34*
	243		0.25	0.50	0.33	0.17	0.09	0.10	0.36	0.05	0.41	0.19	0.17	------	0.37
	271		0.36	0.05	0.22	0.61	0.30	0.40	0.33	0.81	0.37	0.67	0.26	------	0.07
	275		0.39	0.45	0.46	0.22	0.61	0.50	0.31	0.14	0.22	0.15	0.57	------	0.56

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Table S10. Results of the Fisher's method for linkage disequilibrium for each locus pair across all the Iberian populations of flat periwinkles. df is the degree of freedom associated with each test.

Test	Chi Square	*df*	*p-value*
QVOM & VPVX	42.682802	26	0.021*
QVOM & PBL8	39.418933	26	0.044*
VPVX & PBL8	19.240570	26	0.826
QVOM & 193Q	8.554355	18	0.969
VPVX & 193Q	13.984422	16	0.600
PBL8 & 193Q	18.884503	18	0.399
QVOM & KJ2E	20.779466	26	0.753
VPVX & KJ2E	20.941247	26	0.745
PBL8 & KJ2E	31.831491	26	0.199
193Q & KJ2E	8.496889	18	0.970
QVOM & 881	16.752367	26	0.916
VPVX & 881	32.031946	26	0.192
PBL8 & 881	15.374454	26	0.950
193Q & 881	7.861236	18	0.981
KJ2E & 881	11.226815	26	0.995
QVOM & ZIBW	15.410007	22	0.844
VPVX & ZIBW	25.079326	22	0.293
PBL8 & ZIBW	21.273347	22	0.504
193Q & ZIBW	12.133293	16	0.735
KJ2E & ZIBW	25.260369	22	0.285
881 & ZIBW	19.705903	22	0.601
QVOM & 1871	16.218033	22	0.805
VPVX & 1871	31.942201	22	0.078
PBL8 & 1871	12.499775	22	0.946
193Q & 1871	8.037892	16	0.948
KJ2E & 1871	24.636083	22	0.315
881 & 1871	14.804490	22	0.870
ZIBW & 1871	33.027912	22	0.061
QVOM & LHYM	35.786137	26	0.096
VPVX & LHYM	26.067592	26	0.459
PBL8 & LHYM	16.015214	26	0.936
193Q & LHYM	5.978885	18	0.996
KJ2E & LHYM	24.513586	26	0.547
881 & LHYM	10.513195	26	0.997
ZIBW & LHYM	12.832899	22	0.938
1871 & LHYM	22.624161	22	0.423
QVOM & XENN	17.378998	26	0.897
VPVX & XENN	23.991327	26	0.576
PBL8 & XENN	21.275583	26	0.728
193Q & XENN	3.927760	18	1.000
KJ2E & XENN	22.508740	26	0.661
881 & XENN	18.590211	26	0.853
ZIBW & XENN	24.601827	22	0.316
1871 & XENN	7.642228	22	0.998
LHYM & XENN	13.439522	26	0.980
QVOM & 927	27.120927	26	0.403
VPVX & 927	25.927329	26	0.467
PBL8 & 927	19.836576	26	0.799
193Q & 927	17.302246	18	0.502
KJ2E & 927	33.838778	26	0.139
881 & 927	25.181724	26	0.509
ZIBW & 927	11.036169	22	0.974
1871 & 927	6.246889	22	1.000
LHYM & 927	19.346182	26	0.821
XENN & 927	36.930599	26	0.076
QVOM & EVLS	15.143494	20	0.768
VPVX & EVLS	19.594556	20	0.484
PBL8 & EVLS	14.036398	20	0.829
193Q & EVLS	10.361817	12	0.584
KJ2E & EVLS	18.246028	20	0.571

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Table S10. Results of the Fisher's method for linkage disequilibrium for each locus pair across all the Iberian populations of flat periwinkles. df is the degree of freedom associated with each test.

Test	Chi Square	*df*	*p-value*
881 & EVLS	19.921578	20	0.463
ZIBW & EVLS	21.616195	18	0.249
1871 & EVLS	11.408602	18	0.876
LHYM & EVLS	18.493737	20	0.555
XENN & EVLS	16.977363	20	0.654
927 & EVLS	14.127087	20	0.824
QVOM & ZR6M	32.437358	24	0.117
VPVX & ZR6M	16.908286	24	0.853
PBL8 & ZR6M	12.859591	24	0.968
193Q & ZR6M	3.829774	16	0.999
KJ2E & ZR6M	31.707168	24	0.134
881 & ZR6M	24.217351	24	0.449
ZIBW & ZR6M	25.445381	20	0.185
1871 & ZR6M	32.921198	20	0.034*
LHYM & ZR6M	37.801991	24	0.036*
XENN & ZR6M	36.655469	24	0.047*
927 & ZR6M	16.769507	24	0.858
EVLS & ZR6M	14.839452	20	0.786
QVOM & 47	3.956836	12	0.984
VPVX & 47	13.123792	12	0.360
PBL8 & 47	9.610817	12	0.650
193Q & 47	11.267190	12	0.506
KJ2E & 47	4.488477	12	0.973
881 & 47	9.929296	12	0.622
ZIBW & 47	5.722518	10	0.838
1871 & 47	3.881786	10	0.953
LHYM & 47	22.958048	12	0.028*
XENN & 47	11.348853	12	0.499
927 & 47	7.391617	12	0.831
EVLS & 47	0.777934	6	0.993
ZR6M & 47	2.204623	10	0.995
QVOM & TEM7	4.600302	6	0.596
VPVX & TEM7	3.800265	6	0.704
PBL8 & TEM7	1.197180	6	0.977
193Q & TEM7	1.314128	6	0.971
KJ2E & TEM7	3.701901	6	0.717
881 & TEM7	9.235832	6	0.161
ZIBW & TEM7	2.823379	6	0.831
1871 & TEM7	2.160724	6	0.904
LHYM & TEM7	4.508341	6	0.608
XENN & TEM7	2.181209	6	0.902
927 & TEM7	13.675023	6	0.033*
EVLS & TEM7	4.608573	2	0.100
ZR6M & TEM7	0.004024	4	1.000
47 & TEM7	2.630432	6	0.854
QVOM & DAEH	16.738107	24	0.860
VPVX & DAEH	11.495710	24	0.985
PBL8 & DAEH	22.827336	24	0.530
193Q & DAEH	11.571935	16	0.773
KJ2E & DAEH	14.752111	24	0.928
881 & DAEH	20.405318	24	0.674
ZIBW & DAEH	8.824284	18	0.964
1871 & DAEH	19.057168	20	0.518
LHYM & DAEH	16.421803	24	0.872
XENN & DAEH	23.529354	24	0.489
927 & DAEH	29.336374	24	0.208
EVLS & DAEH	9.155282	20	0.981
ZR6M & DAEH	20.788716	24	0.651
47 & DAEH	13.466469	10	0.199
TEM7 & DAEH	4.668977	4	0.323

* indicates significant Hardy-Weinberg deviations before Bonferroni correction ($p<0.05$)
** indicates significant Hardy-Weinberg deviations after Bonferroni correction ($p<0.0002$)

Additional References

Mann, H.; Whitney, D. (1947) On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 18(1), 50-60.

Shapiro, S.; Wilk, M. (1965) An analysis of variance test for normality (complete samples). Biometrika 52, 591-611

Zelditch, M.L.; Swiderski, D.L.; Sheets, H.D.; Fink, W.L. (2004) Geometric morphometrics for biologists: a primer. San Diego, Elsevier Academic Press.